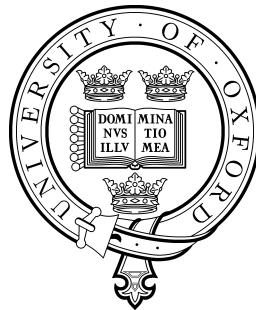


A Bayesian Approach to Financial Model Calibration, Uncertainty Measures and Optimal Hedging



Alok Gupta
Hertford College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2009

This thesis is dedicated to the late
Christine Russell
a wonderful teacher and a constant inspiration,
without whom I would never have come this far.

Acknowledgements

I would like to thank my my joint funding bodies, the Engineering & Physical Sciences Research Council (EPSRC) and Nomura Bank, for their financial support through the CASE award.

I would also like to acknowledge the support shown by the Mathematical Institute of the University of Oxford and the staff of Dartington House in particular. Thanks also goes to Hertford College for looking after me so well during my residence in Oxford. I especially thank Piotr Chruściel for always being available for any concerns I had.

My greatest thanks go to my supervisor Christoph Reisinger. His patience and attention to detail have been unfaltering throughout my DPhil. Christoph has encouraged my research at all times, as well as providing many opportunities for teaching and expanding my academic horizons. He has always had time to go through my problems and, without his knowledge and encouragement, this thesis would not have been.

I would also like to acknowledge the advice and useful conversations I have had with Chris Farmer, Michael Monoyios, Jan Obloj, Mike Giles and Mark Davis. I am especially grateful to Alan Whitley for carefully proof-reading my thesis. I have also been very fortunate to have shared an office with wonderful colleagues, Nikolay, Daniel F, Michael, Andrew, Helen, Daniel S and Jan, who were always willing to discuss my problems, as well as create distractions!

Finally, I would like to thank my family and friends, especially my sisters Abha and Alka and brother Rajeev, for all their care and support. Above all, I am always grateful to my mother and father, without whose love none of this would have been possible. Thanks mum and dad.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Applications	2
1.3	Outline	4
2	Calibration Problem	6
2.1	Introduction	6
2.1.1	The Inverse Problem	7
2.1.2	Well-Posedness	8
2.1.3	Regularisation	9
2.1.4	Bayesian Estimators	10
2.1.5	Advantages of the Bayesian Approach	11
2.1.6	Current Applications of the Bayesian Approach	13
2.2	Consistent Bayes Estimators	14
2.2.1	Literature on Consistent Bayes Estimators	14
2.2.2	Single Price Observation	16
2.2.3	Multiple Price Observations	21
2.2.4	Finite-Dimensional Non-Scalar Parameter	21
2.3	Local Volatility Model Formulation	22
2.3.1	Local Volatility Assumptions	24
2.3.2	Dupire's Formula	24
2.3.3	Error Functional Minimisation Approach	25
2.3.4	Ill-Posedness of Local Volatility Calibration	26
2.3.5	Literature on Local Volatility Calibration	26
2.4	Bayesian Modelling	30
2.4.1	Prior Distribution (Regularisation)	30
2.4.2	Likelihood Function (Calibration Error)	31
2.4.3	Posterior Distribution	32

2.4.4	Choice of Confidence Parameter	32
2.5	Numerical Examples	34
2.5.1	Two Datasets	34
2.5.2	Non-Parametric Discretisation	35
2.5.3	Metropolis Sampling	36
2.5.4	Monitoring Convergence of Metropolis Sampling	38
2.5.5	Results	38
2.5.6	Robustness	46
2.6	Extensions	53
3	Model Uncertainty	56
3.1	Introduction	56
3.1.1	Risk vs Uncertainty	58
3.1.2	Sources of Model Uncertainty	58
3.1.3	Related Research on Model Uncertainty	60
3.1.4	Financial Setup	63
3.1.5	Bayesian Posteriors	64
3.2	Coherent Model Uncertainty Measures	64
3.2.1	Coherent Market Risk Measures	65
3.2.2	Cont's Axioms	66
3.2.3	Worst-Case Measure	67
3.2.4	Motivating New Measures	67
3.2.5	Alternative Axiomatic Approaches	69
3.2.6	New Axioms	70
3.2.7	Representation Theorem for Coherent Model Uncertainty Measures	73
3.2.8	Coherent Measures Using Bayesian Posteriors	74
3.3	Convex Model Uncertainty Measures	76
3.3.1	Convex Market Risk Measures	76
3.3.2	Cont's Axioms	76
3.3.3	Calibration-Error Measure	78
3.3.4	New Axioms	78
3.3.5	Representation Theorem for Convex Model Uncertainty Measures	79
3.3.6	Constructing Penalty Functionals using Bayesian Posteriors	80
3.4	Numerical Examples	81
3.4.1	Robustness	83
3.5	Extensions	85

4 Optimal Hedging	87
4.1 Introduction	87
4.1.1 Delta Hedging using Deterministic Volatility Functions	88
4.1.2 Hedging Volatility and Jump Risk	90
4.2 Hedging in the Presence of Model Uncertainty	91
4.2.1 Literature on the Impact of Model Uncertainty on Hedging	91
4.2.2 Hedging Formulation	93
4.2.3 Motivating Examples	95
4.2.4 Previous Bayesian Approaches to Hedging	96
4.2.5 Bayesian Hedging using Loss Functions	97
4.3 Hedging Error Loss Functions	99
4.3.1 Hedging Improvement	101
4.3.2 Derived Model Uncertainty Measures	102
4.3.3 Relationship with Utility Functions	104
4.4 Numerical Examples	105
4.4.1 Heston Model	106
4.4.2 Dynamic Hedging Strategies	108
4.4.3 Computation of the Loss Functions	110
4.4.4 Bayesian Updating (Recalibration)	110
4.4.5 Hedging Results	112
4.4.6 Robustness	118
4.5 Extensions	123
5 Conclusions	125
5.1 Summary of Main Results	125
5.2 Interpretation of Findings	126
5.3 Recommendations for Future Research	128
A Datasets	131
A.1 Calibration Prices	131
A.2 Calibration Constants	132
B Algorithms	133
B.1 Sobolev Norm Induced Inverse Covariance Matrix	133
B.2 Fast Fourier Transform Formula for Pricing in the Heston Model	134

C Proofs	136
C.1 Dupire's Formula	136
C.2 Terminal Hedging Error for Discrete Delta-Neutral Portfolio	137
C.3 Terminal Hedging Error for Discrete Delta-Vega-Neutral Portfolio	138
References	139

List of Figures

2.1	For the simulated dataset: plot of calibrated surfaces for simulated prices	39
2.2	For the simulated dataset: 95% and 68% confidence intervals for volatility of simulated paths	40
2.3	For simulated dataset: prices for barrier call option	42
2.4	For simulated dataset: prices for American put option	42
2.5	For the simulated dataset: consistency of the Bayesian estimate of the local volatility surface	43
2.6	For S&P 500 dataset: plot of calibrated surfaces	45
2.7	For S&P 500 dataset: prices for American put option	46
2.8	Distance between original posterior probability measures and posterior probability measure for different priors	48
2.9	Robustness of barrier call prices to changes in the prior	49
2.10	Robustness of American put prices to changes in the prior	49
2.11	Distance between original posterior probability measures and posterior probability measure for different variations in the observed data	50
2.12	Robustness of barrier call prices to changes in the observed data	51
2.13	Robustness of American put prices to changes in the observed data	52
3.1	Motivating example for extension of model uncertainty measures	68
3.2	For simulated dataset: example model uncertainty values for 5 options	82
3.3	For simulated dataset: histograms of prices calculated in different models	82
3.4	For S&P 500 dataset: model uncertainty values for 5 options	83
3.5	Robustness of model uncertainty values to changes in the prior	84
3.6	Robustness of model uncertainty values to changes in the observed data	84
4.1	1 st Motivating example of performance of MAP and Bayesian hedging strategies . .	95
4.2	2 nd Motivating example of performance of MAP and Bayesian hedging strategies . .	96
4.3	For local volatility underlying: Bayesian hedging improvements for European call . .	113
4.4	For local volatility underlying: Bayesian hedging improvements for barrier option . .	114

4.5	For Heston underlying: Bayesian hedging improvements for European call	116
4.6	For Heston underlying: Bayesian hedging improvements for barrier option	117
4.7	Robustness of Bayesian hedging improvements for European call to changes in the prior	119
4.8	Robustness of Bayesian hedging improvements for barrier option to changes in the prior	120
4.9	Robustness of Bayesian hedging improvements for European call to changes in the observed data	121
4.10	Robustness of Bayesian hedging improvements for barrier option to changes in the observed data	122

List of Tables

2.1	For the simulated dataset: convergence Check (PSRF Values)	39
2.2	For S&P 500 dataset: convergence Check (PSRF Values) for the S&P 500 dataset . .	45
A.1	For the simulated dataset: European call prices (units of 10^3)	131
A.2	For the S&P 500 dataset: European call prices (\$)	131
A.3	Numerical examples constants for calibration process	132

Chapter 1

Introduction

1.1 Motivation

Since the model proposed by Black and Scholes in their seminal 1973 paper [17], the variety of financial models has grown and grown. With this growth have come more complex and sophisticated models trying to capture the behaviour of underlying market processes. Today we have stochastic volatility models for equity processes and numerous jump processes for credit models, to mention but a couple. Typically, an agent will want to use her model to price or hedge an instrument in the market. But before she can do this she must first mark her model to market — that is, *calibrate* her model to observable prices.

Most commonly used calibration instruments are the simplest such as European calls and puts. Calibration is necessary to avoid introducing arbitrage into the market and making the agent vulnerable to other agents making riskless profits from the first agent's incorrect prices. However, calibration is not a straightforward process. In the original Black-Scholes model there is one scalar volatility parameter that has to be estimated but, in most of the commonly used models today, sometimes entire functions have to be calibrated. This is not an easy task.

Firstly, there can be inconsistencies and/or mis-pricings in the prices of the calibration instruments themselves, so that no parameter could correctly reproduce all the observed prices. Alternatively, there might not be enough observable market prices to uniquely determine the calibration parameter. Thirdly, there is often no guarantee of the stability of the calibrated model — a small change in the market prices may lead to a large and disproportionate change in the calibration parameter of the model. This becomes particularly hazardous when we try to compute the corresponding hedging strategies.

Despite these pitfalls, much of the literature on calibration methods still focuses on finding the single best-fit parameter for a given set of market prices. Little attention is paid to measuring how robust this parameter is or whether other parameters can reproduce market prices equally/almost as well. This is a major shortcoming in much of the current literature. A measure of the uncertainty of

the calibration parameter is vital for two reasons. Firstly, it gives a good indication of how suitable our model and calibration dataset are. Secondly, it enables better risk management and provides clear quantitative measures of potential pricing errors, hedging losses, and other important financial quantities. For example, greater uncertainty of the exact value of the calibration parameter will imply a greater degree of (model) risk in any investments made based on the subsequent predictions of our calibrated model.

Hence, in this thesis we recast the calibration problem into a Bayesian framework so that, instead of only finding a best-fit point estimate for the calibration parameter (which may be a scalar, a finite-dimensional vector, or a function), we can actually extract an entire distribution of calibrated parameters. We can then use this distribution in the ways outlined above to decide how suitable a model is, give a measure of the model uncertainty and construct optimal hedging strategies. The Bayesian method is a very natural way of assigning probabilities to competing candidate models. Moreover, it is a rigorous and practical method for incorporating all qualitative and quantitative information available. The Bayesian method is particularly suited to the field of mathematical finance in which new data (market prices) are observed daily and can be used to update Bayesian estimates. Further detail on the advantages of the Bayesian philosophy is provided in Section 2.1.5.

1.2 Applications

The applications of constructing a distribution of calibration parameters are broad and far-reaching. As outlined, in this thesis we hope to find a whole family of candidate calibration parameters, each with a corresponding probability of being the true value (for any given day or time period). For a trader this wealth of extra information can, for example, give her 95% or 68% confidence intervals for the price of an American option or barrier option. Instead of a single point estimate price, the trader sees a range of prices, with varying degrees of probability. This is further detailed in Chapter 2.

Behind the front desk, for risk managers say, the extra information gives them a precise and quantitative measure of how right or wrong the traded prices could be. The distribution of calibrated parameters will enable forecasting confidence intervals and a true idea of a portfolio's possible positions. The end game of course is to investigate whether or not the choice of model type can reduce the risk and tighten bounds on confidence intervals.

Furthermore, as we detail in Chapter 3, the posterior distribution can be used to construct practical numerical model uncertainty measures to enable comparison of alternative financial investments and the money value of the model uncertainty associated with these investments. Equally, if not more, pertinent, is that the posterior distribution is used in Chapter 4 to construct optimal hedging strategies that maximise expected hedging profits or minimise expected absolute deviations of hedging errors.

It is important to remark here that the approach of this work is very general and can in principle be applied to the calibration problem for any type of underlying process: diffusion, jump-diffusion, Lévy, and others. Moreover, for each process we can try to calibrate any model of choice. In the case of diffusion, this might be the generalised semi-parametric driving processes proposed by Bingham & Kiesel [15], or the stochastic volatility model proposed by Hull & White [88], or the structural model of Merton [107] in the context of credit defaults. Equally, if studying jump diffusion processes we might calibrate the model looked at by Zhou [135] or in the case of Lévy processes the model considered by Hilberink & Rogers [83].

In this thesis, we use the local volatility model as a working example throughout to clarify and demonstrate our method and results. In the local volatility model the volatility of an asset price process is assumed to be a function of the asset price and time; full detail of the model is provided in Section 2.3. We choose the local volatility model for a number of reasons.

Firstly, it is a conceptually simple model — a deterministic function of only two variables (time and asset price) — for which a satisfactory practical calibration method has not yet been found. We describe the difficulties in more detail in the next chapter. But in essence, it is an underdetermined problem, because there are too few prices to calibrate the surface to; and these prices are not perfectly accurate either. Thus the local volatility is a good model to exemplify the mathematical rigour of the Bayesian framework.

Secondly, the local volatility model is a hugely difficult model to calibrate because it is a continuous function and hence chosen from an infinite-dimensional space. Compared to a Hull-White or Heston stochastic volatility model in which there are only a few scalar parameters to calibrate, the local volatility model, is far more demanding because we try to calibrate a two dimensional function. By demonstrating the Bayesian method to solve the more difficult non-parametric calibration problem, it should be immediately clear how to apply the method to the simpler parametric calibration problems.

Thirdly, the local volatility model, although widely used in practice, is not a satisfactory model, and this makes it a compelling candidate for testing the robustness of our Bayesian methods. The local volatility model is popular amongst traders because it can be perfectly fitted to market volatility smiles and the model is complete [39]. However, compared with stochastic volatility models and Lévy models, the local volatility is dynamically inconsistent. It is inconsistent in the sense that it tells us today what the volatility will be in the future for a given asset price, but when tomorrow we calibrate to tomorrow's observed option prices the surface changes and gives new estimations. Hence, if dynamic Bayesian hedging can be shown to raise the hedging profits resulting from using the local volatility model, as we do in Chapter 4, then it is a good indication that our methodology will improve the performance of consistent models.

It is for all these reasons that the local volatility is thus an ideal model to use throughout as an example for our non-parametric Bayesian techniques. We recall here that the local volatility model by definition assumes market completeness, i.e. the payoff of any contract can be perfectly replicated using traded assets. In Chapter 4 we drop this assumption when we carry out numerical examples using the Heston stochastic volatility model.

1.3 Outline

This thesis aims to make contributions to the practical implementation of Bayesian methods for calibrating non-parametric financial models, constructing model uncertainty measures and finding optimal hedging strategies. The thesis takes the format described in the following few paragraphs.

In the next chapter, Chapter 2, we tackle the calibration problem. We start by reformulating the problem in a classical inverse problem setup and review the literature for inverse problems. We then detail the Bayesian approach to solving inverse problems and prove some consistency theorems for Bayesian estimates. This chapter contributes new consistency results for Bayes estimators for non-identically distributed observations. In the third section of Chapter 2 we consider the local volatility model as a case study to implement the aforementioned approaches. A practical method for constructing the local volatility surface calibration problem is given and some numerical examples are worked through to show how to find calibrated surfaces. The final section of Chapter 2 suggests generalisations to other models and further extensions.

In Chapter 3 we consider model uncertainty measures and review current contributions to the topic. In particular we study two classes of model uncertainty measures: coherent and convex. This is motivated by the popular coherent and convex market risk measures introduced in the papers by Artz et al. [2] and Frittelli & Gianin [69] respectively earlier this decade. The contributions of Chapter 3 are the design of new axiomatic frameworks for each class of model uncertainty measure and proof of a representation theorem for each class. This then enables efficient construction of a wide variety of coherent and convex measures. In both cases we demonstrate how to use the Bayesian posteriors constructed in Chapter 2 to build meaningful model uncertainty measures. In the penultimate section of Chapter 3 we return to the local volatility numerical examples and compute the convex model uncertainty values of some options. The final section recommends some extensions to the theory of model uncertainty measures.

Chapter 4 looks to use the Bayesian posteriors of Chapter 2 to deal with the model uncertainty values seen in Chapter 3. We consider how best to use all the available information, post the calibration procedure, to design optimal hedging strategies. The first section of Chapter 4 outlines why correct hedging is important and reviews literature on different hedging techniques. The major contribution of this chapter is the design of optimal hedges using Bayesian loss functions. Instruction is provided on how to construct these loss functions and we prove some conditions necessary in order

to use these loss functions to create new convex model uncertainty measures satisfying the axioms presented in Chapter 3. Furthermore, these conditions are shown to be identical to those necessary to achieve an equivalence between loss functions and utility functions. The penultimate section provides practical guidance on calculation of the dynamic hedging strategies and some numerical examples in which we compare the Bayesian hedging strategies with popular delta and delta-vega hedging strategies. The final section of Chapter 4 demonstrates how the loss functions can be customised to suit an investor's preferences.

The final chapter, Chapter 5, summarises the chief results of the thesis, interprets these findings and suggests possible avenues for future investigation. The appendices and references follow.

Chapter 2

Calibration Problem

In this chapter we introduce the general calibration problem in financial models and review current literature on the subject. We reformulate the problem into a Bayesian framework to attain posterior distributions for calibration parameters. We give conditions on the value function under which the corresponding Bayes estimator is consistent. Finally, we apply our results to a discretised local volatility model and work through numerical examples to clarify the construction of Bayesian posteriors and its uses.

2.1 Introduction

As mentioned in Chapter 1 the growth in financial products has been vast in the previous decades. And this growth has motivated ever more sophisticated and realistic financial models. However, before theoretical models can be used to make accurate inferences about financial products or strategies, they must be calibrated so that they reproduce existing market values. But calibration is not usually straightforward. In practice it is unclear whether perfect calibration is even possible. In reality the calibration instruments we try to mark our model to are only observable in the market up to some *bid-ask spread*, which is the interval of values between what an agent is willing to pay for the instrument (at the lower end) and what an agent is willing to sell the instrument for (at the upper end). Hence, it is standard practice to find the model that minimises the calibration error — the sum of the absolute differences between the calibration prices (or mean of the bid and ask prices) and the corresponding prices produced by the model.

There is a wealth of literature and techniques proposed for calibrating models. Towards the end of this chapter we consider the local volatility model in particular. Authors such as Jackson et al. [89], Lagnado & Osher [98] and Coleman et al. [28] have all looked at minimisation techniques and regularisation functions for finding the ‘best-fit’ local volatility surface. Further analysis of these methods and their improvement has been detailed by Chiarella et al. [26]. However, the literature lacks substantial investigation on robustness. In a nutshell, the process of calibration can be classed as an *inverse problem*: we have a parametrised model to price instruments; we then

observe some prices and try to find the model parameter which gives those prices. It is called an inverse problem because we explicitly know how to get from the model to the prices but not from the prices to the model. Furthermore, as we shall detail later, the problem is often ill-posed because it is underdetermined and non-robust to small changes in the observed prices.

Hence, the approach of this thesis is to no longer focus on finding a *best-fit* solution, but on finding an entire *distribution* of solutions. So instead of selecting the model which best replicates prices, we find all models which replicate prices to within a pre-specified tolerance level of error. This is an approach that recently Hamida & Cont [79] have adopted as part of their investigation into model uncertainty. Where this work differs from [79] is by recasting the problem into a Bayesian framework. The approach allows us to assess the stability of the estimators by observing how the Bayesian posterior distribution varies with small changes in the prior and calibration prices.

In this chapter we provide a practical method for constructing prior and likelihood functions and spend some time analysing the long run properties of Bayesian posteriors. The chapter is divided as follows. In the next few sections we formalise the general calibration problem, common regularisation approaches and the Bayesian approach. In Section 2.2 we prove some important results on the consistency of Bayes estimators in general models. In Section 2.3 we introduce the local volatility model and review literature on the methods for its calibration. In Section 2.4 we detail a method for constructing Bayesian prior and likelihood functions. In Section 2.5 we explain the computational algorithms and present some numerical examples. We also provide quantitative details of robustness of the scheme. Section 2.6 discusses some possible future extensions to the work of the chapter.

2.1.1 The Inverse Problem

Suppose we observe a price process $S = (S_t)_{t \geq 0}$ and model it as a function of time t , some stochastic process(es) $Z = (Z_t)_{t \geq 0}$, and parameter $\theta \in \Theta$, i.e.

$$S_t = S(S_0, t, (Z_u)_{0 \leq u \leq t}; \theta) \quad (2.1)$$

by abuse of notation of S and where S_0 is the time-0 value of the price process S . Let $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ be the filtration on the set of scenarios Ω generated by Z , so S is an \mathcal{F} -adapted process.

Remark 2.1.1. Although we are specifically interested in non-parametric financial models and thus, for example, when θ represents a local volatility function, we will assume henceforth that θ is a finite-dimensional vector: $\Theta \subseteq \mathbb{R}^M$ for $M \in \mathbb{N}$. This is a sensible approximation to make, since in practice we will have to discretise any function to be able to estimate it using numerical methods. However, we remark that the nature of the problem changes in the transition from finite- to infinite-dimensional θ .

Now consider an option over a finite time horizon $[0, T]$ written on $S(\theta)$ and with payoff function h . Let the time t value of this option be written as $f_t(\theta)$, where we include the argument θ to emphasise the dependence of this price on the model parameter θ . By no-arbitrage assumptions we have explicitly,

$$f_t(\theta) = \mathbb{E}^{\mathbb{Q}}[B(t, T)h(S(\theta))|\mathcal{F}_t]$$

with respect to some the risk-neutral measure \mathbb{Q} and where $B(t, T)$ is the discount factor over the interval $[t, T]$, possibly stochastic.

Remark 2.1.2. For any particular θ , $f_t(\theta)$ is an \mathcal{F} -adapted process. That is, fixing θ , $f_s(\theta)$ is known for all $s \leq t$ but $f_s(\theta)$ is a random variable for all $s > t$.

Suppose at time $t \in [0, T]$ we observe a set of such option prices $\{f_t^{(i)}(\theta) : i \in I_t\}$, possibly with noise $\{e_t^{(i)} : i \in I_t\}$ where I_t is an index set of data. In other words, we observe

$$V_t^{(i)} = f_t^{(i)}(\theta^*) + e_t^{(i)} \quad (2.2)$$

for $i \in I_t$ and where θ^* is the true parameter. Then the calibration problem is to find the value of θ that *best* reproduces the observed prices $\{V_t^{(i)} : i \in I_t, t \in \Upsilon_n([0, T])\}$, whatever *best* means. Here $\Upsilon_n([0, T]) = \{t_1, \dots, t_n : 0 = t_1 < t_2 < \dots < t_n \leq T\}$ defines a partition of the interval $[0, T]$ into n parts.

However, before attempting to find the solution θ it is first necessary to decide whether a stable solution exists at all.

2.1.2 Well-Posedness

We call a mathematical problem *well-posed* if it satisfies Hadamard's criteria (see for example [59]):

- i) For all admissible data, a solution exists.
- ii) For all admissible data, the solution is unique.
- iii) The solution depends continuously on the data.

If on the other hand a mathematical problem violates one or more of the above criteria then we call it *ill-posed*. Parameter identification problems are often ill-posed. In the context of calibration, we start by assuming we can find a solution fitting the data to within an acceptable error tolerance, δ say, and hence satisfying i). However, we cannot guarantee properties ii) and iii). The effects of violating either of these two properties will be seen for pricing and hedging. Furthermore, the admissible data is almost always noisy — the values are only observed with added error as in (2.2) — so we assume the true values to lie within some confidence interval around the observed value. In our context, prices are never observed exactly but only to within a bid-ask spread, with the true

price (usually) lying somewhere in this spread of prices. So the bid-ask spread can be thought of as this error confidence interval.

If there is more than one possible solution, i.e. more than one calibrated parameter, then we call the inverse problem *underdetermined*. This happens when we do not have enough market prices to restrict the value of the calibrated parameter. In this situation, choosing the wrong calibrated parameter will lead to incorrect pricing and hedging of other options, which can result in losses for a trading agent. Equivalently, if a solution does not depend continuously on the data, i.e. market prices, then a small mis-pricing in the market of one of the observed prices can lead to a disproportionately large error in the chosen calibrated parameter and again incorrect pricing and hedging. This problem is usually the more serious and difficult to overcome.

2.1.3 Regularisation

We call the process of approximating an ill-posed problem by a well-posed problem *regularisation*. A vast literature (for example [59] and [130]) exists on handling ill-posed problems and especially ill-posed inverse problems.

Let us consider a general inverse problem in which we know the forward function f and want to solve

$$f(\theta) = V \quad \theta \in \Theta, V \in \mathcal{V} \quad (2.3)$$

for finite-dimensional θ , but do not know the inverse function f^{-1} . Suppose further that we can only observe an approximation V^δ for V , $\|V^\delta - V\|_{\mathcal{V}} \leq \delta$, and are instead trying to solve $f^{-1}(V^\delta) = \theta^\delta$. Assume that f^{-1} does not satisfy Hadamard's condition ii) and/or iii) from the previous section. Then one way to regularise the problem is to restrict possible solutions to a subset $\hat{\Theta} \subset \Theta$ where, for example, $\hat{\Theta}$ might be chosen to be a compact set so that the problem becomes stable under small changes in V .

The more widely used approach however is to replace f^{-1} with a *regularisation operator* f_λ^{-1} with *regularisation parameter* $\lambda > 0$ which depends on δ and/or V^δ . The operator and parameter are chosen so that

$$\begin{aligned} \lambda &= \lambda(\delta, V^\delta) > 0, \\ f_\lambda^{-1} : \mathcal{V} &\rightarrow \Theta \text{ is bounded } \forall \lambda \in (0, \lambda_0), \\ \lim_{\lambda \rightarrow 0} \sup \{ \|f_\lambda^{-1}(V^\delta) - f^{-1}(V^\delta)\|_{\Theta}\} &= 0. \end{aligned}$$

This ensures that $(f_\lambda^{-1}(V^\delta)) =: \theta_\lambda^\delta \rightarrow \theta^\delta$ as $\lambda \rightarrow 0$. The regularisation operator aims to make the function to be minimised convex so that the solution is unique and easily located, by a conjugate-gradient search algorithm for example.

Because it is not always practical to exactly solve $f_\lambda^{-1}(V^\delta) = \theta^\delta$, we instead form $g_\lambda^\delta(\theta^\delta) = f_\lambda(\theta^\delta) - V^\delta$ and solve

$$\text{find the } \theta^\delta \text{ which minimises } \|g_\lambda^\delta(\theta^\delta)\|_{\mathcal{V}}.$$

It still remains however to find a regularisation operator and parameter. There are several methods for doing so (see [130] for details): using the spectrum of operator f , using Fourier, Laplace, and other integral transformations. However, the most obvious and common way to construct g_λ^δ is by adding a *stabilising function* $h : \Theta \rightarrow \mathbb{R}$ to the original function g^δ , so the regularisation operator becomes

$$g_\lambda^\delta = g^\delta + \lambda h. \quad (2.4)$$

Hence our original problem (2.3) becomes

$$\text{find the } \theta^\delta \text{ which minimises } \|g^\delta(\theta) + \lambda h(\theta)\|_{\mathcal{V}}. \quad (2.5)$$

An appropriate choice for h varies from problem to problem, but common practice is to take a *Tikhonov functional* [130]. For $\theta = \theta(t)$, Tikhonov functionals of the p th order are given by

$$h(\theta) = \int_{t=0}^{t=T} \sum_{r=0}^p a_r(t) \left(\frac{d^r \theta}{dt^r} \right)^2 dt \quad (2.6)$$

for some coefficient functions $a_r(t)$, usually taken to be constant. It is worth observing that the functional (2.6) is simply the natural weighted *Sobolev norm* associated with the *Sobolev space* W_2^p . As mentioned, one way to force uniqueness of solution is to restrict the solution space Θ to suitable $\hat{\Theta}$. The Tikhonov functional works in a similar way, favouring solutions which belong to the subset of functions with small Sobolev norm.

2.1.4 Bayesian Estimators

Bayesian theory can be used to estimate the value of an unknown parameter. It provides a rigorous framework for combining prior information with observations to calculate likely values. Suppose we wish to estimate the value of some (finite-dimensional) parameter θ . Assume we have some prior information for θ (for example that it belongs to a particular space, or is positive, or represents a smooth function), summarised by a *prior* density $p(\theta)$ for θ . And suppose we observe some noisy data $V = \{V_t : t \in \Upsilon_n\}$:

$$V_t = f_t(\theta^*) + e_t$$

for all $t \in \Upsilon_n$, where θ^* is the true parameter, e_t is some random noise and Υ_n is an index set of size n . Note that this is a special case of (2.2) with one observation per time t , i.e. $|I_t| = 1$ for all t . In what follows, by abuse of notation, the function p will depend upon its argument. Then $p(V|\theta)$ is the probability of observing the data V given θ , and is called the *likelihood* function.

Now, an application of Bayes rule (see for example [100]) implies that the *posterior* density of θ is given by

$$p(\theta | V) = \frac{p(V|\theta) p(\theta)}{p(V)},$$

where $p(V)$ is given by

$$p(V) = \int p(V|\theta) p(\theta) d\theta \quad (2.7)$$

and it is assumed that $\theta \in \mathbb{R}^M$ for finite-dimensional M .

Definition 2.1.3. A function $L : \mathbb{R}^{2M} \rightarrow \mathbb{R}$ is a *loss function* iff

$$\begin{cases} L(\theta, \theta') = 0 & \text{if } \theta' = \theta \\ L(\theta, \theta') > 0 & \text{if } \theta' \neq \theta, \end{cases}$$

where $\theta, \theta' \in \mathbb{R}^M$

Assumption 2.1.4. Since the loss function should penalise estimators which are further from the true value more than those which are closer, it is usual to assume that the larger the error, the greater the loss. So we assume L is a (not necessarily strictly) increasing function of $\|\theta - \theta'\|$ for some norm $\|\cdot\|$ on \mathbb{R}^M .

Definition 2.1.5. Given data V and loss function L , a corresponding *Bayes estimator* $\theta_L(V)$ is a value of θ which minimises the expected loss with respect to the posterior, i.e.

$$\theta_L(V) = \arg \min_{\theta'} \left\{ \int L(\theta, \theta') p(\theta|V) d\theta \right\}.$$

Remark 2.1.6. The minimiser $\theta_L(V)$ is not necessarily unique.

Example 2.1.7. Different choices of loss function give well-known Bayes estimators:

1. $L_1(\theta, \theta') = \|\theta - \theta'\|_2^2$ gives Bayes estimator $\bar{\theta} = \theta_{L_1}(V) = \mathbb{E}[\theta|V]$, which is the *mean* value of θ with respect to the Bayesian posterior density $p(\theta|V)$.
2. $L_2(\theta, \theta') = 1 - 1_{\theta=\theta'}$, known as the uniform loss distribution, gives the Bayes estimator $\theta_{MAP} = \theta_{L_2}(V) = \max\{p(\theta|V)\}$, which is the *maximum a posteriori* (MAP) estimator; it is the value which maximises the posterior density.

2.1.5 Advantages of the Bayesian Approach

For full details on Bayesian data analysis and applications the reader is referred to [70], [11] and [126]. We highlight some of the main arguments in this below.

As a method for solving inverse problems, the Bayesian framework offers some advantages over the regularisation method discussed in Section 2.1.3. Point estimates such as $\theta_{L_1}(V)$ and $\theta_{L_2}(V)$ given in the previous section are useful, but meaningless without some measure of their correctness. The Bayesian approach offers a formal and consistent way to attach confidence to estimates. Equally,

the approach provides a coherent way to incorporate *all* available information regarding the unknown parameter, clearly differentiating between the a priori and observed information.

Later, in Section 2.4.3, we see how, with special choices for the prior and likelihood, we recover the regularisation operator (2.4) and the MAP estimator is equivalent to the solution of (2.5). However, the advantage of the Bayesian approach is that we also discover a *natural* value for the regularisation parameter λ , as we explain in Section 2.4.4. This is important because in the regularisation method this is something that is largely found through trial and error. The choice of stabilising term is often ad hoc or non-rigorous and therefore unsatisfactory. In the Bayesian framework, however, each term is meaningful and non-arbitrary.

Opponents of the Bayesian approach to data analysis often argue that it is fundamentally wrong to treat an unknown model parameter as a random variable and attach a distribution to it. They argue that the model parameter is unknown but not random. The counter argument is that in some cases it is as important to be able to measure the uncertainty of a model parameter as it is to find the model parameter. One method of measuring the potential error is precisely to put a distribution on the model parameter and regard it as a random variable. A second argument against the use of Bayesian theory is that the prior is inappropriate and meaningless. The argument is that scientists should not analyse data with any preconceptions or bias. However, in the mathematics of this thesis, the prior is a powerful method of formally incorporating underlying assumptions. In mathematical finance, ideas such as *no arbitrage* and *market completeness* and *perfect knowledge* are fundamental to the subject and have to be used. The Bayesian prior provides a neat method for doing so. For example, no-arbitrage assumptions can be incorporated into the prior by attaching zero prior probability to parameters which introduce arbitrage opportunities for calibration instruments.

Other, more practically minded opponents of the Bayesian methodology sometimes argue that the assignment of probabilities to different parameters is too arbitrary, subjective and difficult. For example, Cont [30] argues that assigning weights to models ‘requires too much probabilistic sophistication on the part of the end user’. However, the view of this thesis is that it is in fact very natural to specify prior weights on candidate parameters. It is precisely because the problem is ill-posed and we must involve some regularisation technique that a prior naturally and unavoidably arises (see Remark 2.4.1). Whether we choose to call the regularisation adjustment a roughness penalty function or smoothing term or prior is, in the opinion of the author, a preference more of terminology than philosophy. Moreover, it is not even important that the prior should be very accurate or very carefully deliberated over. The Bayesian method is robust to inaccuracies and mis-specifications of the prior. This is demonstrated for our particular numerical examples in Section 2.5.6.

These arguments aside, once the Bayesian posterior has been found a variety of useful analyses can be performed:

- *credible sets* (also known as confidence intervals) can be generated by approximating the local behaviour of the posterior at a maximum (global or local) by a Gaussian distribution. For example, if the approximation about θ_0 has standard deviation s then a 68% confidence interval would be given by $[\theta_0 - s, \theta_0 + s]$ (assuming a Gaussian distribution about θ_0). Even without making a Gaussian approximation, one can find, for example, an interval containing 95% of the posterior distribution.
- *marginal distributions* of a component of θ can be found by integrating the joint posterior with respect to the other components. Viewing the marginal distribution of each component is useful in understanding how sensitive the joint posterior is to each of the components of θ and also how much each component can vary.
- *inferences* can be made about another quantity of interest, W say, that is a function of θ . The spread of W can be measured and hence the errors associated with using a single point estimate for θ can be calculated.

2.1.6 Current Applications of the Bayesian Approach

The application of Bayesian theory to calibration problems in mathematical finance, although not a novel idea, is something that has only gathered weight over the previous two decades. In the early 1990s Jacquier et al. [91] showed that Bayes estimators for a particular class of stochastic volatility models outperform the widely used method of moments and quasi-maximum likelihood estimators. More recently, Bhar et al. [13] and Para & Reisinger [113] have considered dynamic Bayesian approaches to calibrating instantaneous spot and forward interest rates respectively.

However, current attention has turned to using the Bayesian framework to examine the implications of parameter uncertainty in financial models. Jobert, Platania and Rogers [92] consider a Bayesian approach to explain the consistently large observed excess return earned by risky securities over the return on T-bills. They argue that, by dropping the assumption that the parameters of the dividend process are known to an agent but instead the agent only has some prior beliefs of these parameter, the excess rates of return are a natural consequence. Similarly, Monoyios [108] examines the effects of drift parameter uncertainty in an incomplete market in which claims on non-traded assets are optimally hedged by a correlated traded asset. Using Bayesian learning, [108] concludes that terminal hedging errors are often very large. Jacquier & Jarrow [90] look at the effect on parameter uncertainty and model error in the Black-Scholes framework. They use Bayes estimators to infer values for option prices and hedge ratios and assess non-normality of the posterior distributions.

Closer to the example of the local volatility model used in this chapter are the works by Darsinos & Satchell [40], [41]. The first paper, [40], formulates a joint prior for the asset price S_t and the Black-Scholes implied constant volatility σ using historical log-returns of the asset price. The prior

is updated using newly observed returns to give the posterior. The posterior is then transformed to a function of the asset price S_t and Black-Scholes European call price c and marginalised to give the probability density function for the option price c . The second paper, [41], uses this density to forecast European call option prices one day ahead and numerical experiments show substantial improvement to benchmark mean implied volatility procedures, especially in terms of hedging profits.

2.2 Consistent Bayes Estimators

In order to judge the validity of an estimator, we would like it to satisfy certain properties. For example, we might expect the estimator to be unbiased. Or instead we might hope that, as we observe more data V , the estimator converges in some sense to the true value. Technical definitions for these concepts are set out below.

Definition 2.2.1. (Adapted from [100]). A sequence of estimators $\hat{\theta}_n = \hat{\theta}(\{V_t : t \in \Upsilon_n, |\Upsilon_n| = n\})$ of an unknown θ is said to be *consistent* if $\hat{\theta}_n$ tends in probability to θ (written $\hat{\theta}_n \xrightarrow{P} \theta$) for all $\theta \in \Omega$, i.e.

$$\forall \theta \in \Omega \quad \forall \delta > 0 \quad \mathbb{P}_\theta[|\hat{\theta}_n - \theta| \geq \delta] \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

where \mathbb{P}_θ denotes the probability measure corresponding to θ .

Definition 2.2.1 essentially states that, as more data is collected, knowledge (the Bayes estimator) of the unknown parameter should become more accurate and precise. As Ghosal [72] points out in his review of consistency, this property is crucially important since the violation of consistency puts serious doubts against inferences based on the inconsistent posterior distribution. Secondly, as Wasserman [131] remarks, we should also desire that the Bayesian posterior is not dominated and led astray by the priors. For these reasons we try to prove consistency of the Bayes estimators which, in our case, are for observations that are independent but not identically distributed. And in order to prove consistency of these Bayes estimators in the next sections we will use the following definition.

Definition 2.2.2. The *Dirac* measure \mathbb{P}_{δ_a} is an ‘atomic’ measure given by $\mathbb{P}_{\delta_a}[\{a\}] = 1$.

With reference to (2.1) the unknown parameter is assumed to be θ (scalar and constant for Sections 2.2.2 and 2.2.3).

2.2.1 Literature on Consistent Bayes Estimators

We briefly review here some of the existing research and results on the asymptotics of estimates found using the Bayesian method.

The majority of the literature on the consistency of Bayes estimators assumes observations are independent and identically distributed (i.i.d.). The first major result on the consistency of Bayes

estimators was provided by Doob [49] in 1949, who showed that, for every prior distribution $p(\theta)$ on parameter space Θ and convex loss functions, the Bayes estimators are consistent, except possibly for values of θ belonging to null sets with respect to the prior. Some years later Doob's result was extended by Schwartz [122] to loss functions that are not convex. More recently, Fitzpatrick [62] has observed the relationship between Tikhonov regularisation and Bayesian maximum likelihood estimators in the Gaussian noise framework and derived consistency results for the latter, but again with respect to i.i.d. observations.

However, borrowing an example from Wasserman [131], we see how Doob's result can be misleading. Suppose Y_1, \dots, Y_n are normally distributed $N(\theta, 1)$ and the prior $p(\theta) = \delta(\theta)$ i.e. is a point mass at zero. Then the posterior will also be a point mass at zero. So, except for the case when $\theta = 0$, the posterior will not accumulate around the true value and the Bayes estimator will be inconsistent. However, if we denote by A the event that the posterior eventually accumulates around the true value of θ , then Doob's result concludes that A has probability one. And this seems counterintuitive to the fact that the posterior is inconsistent. The cause of this paradox is that [49] uses the Bayesian's own prior $p(\theta)$ when it concludes A has probability one. For, if $\theta = 0$ then the Bayes estimator is consistent and $\theta = 0$ is a sure event according to $p(\theta) = \delta(\theta)$. Clearly Doob's consistency fails on a null set, but in this example the null set was not very large. Although usually less of a problem for one- or two-dimensional parameters θ , for higher and infinite dimensions, it is likely there will be some very large null sets.

The above pitfall has been stressed by Diaconis & Freedman [48]. Their study shows that mechanical Bayesian non-parametric techniques, such as assuming a Dirichlet prior, can lead to inconsistent Bayes estimates. The Dirichlet distribution of order $M + 1 \geq 2$ with parameters $\alpha_1, \dots, \alpha_{M+1} \geq 0$ has probability density function on \mathbb{R}^M given by

$$f(\theta_1, \dots, \theta_{M+1}; \alpha_1, \dots, \alpha_{M+1}) \propto \prod_{i=1}^{M+1} \theta_i^{\alpha_i - 1}$$

for all positive $\theta_1, \dots, \theta_{M+1}$ summing to unity. [48] notes that, especially in high dimensions, large null sets in the prior can swamp the observed data, no matter how much is collected, and lead to inconsistent Bayes estimates. Wasserman [131] builds upon these results to show that, in some examples, Bayesian posteriors on infinite-dimensional spaces do not give consistent estimators and that it is more appropriate to use frequentist procedures instead. Recall the frequentist point of view is that the probability of an event is the limit of its relative frequency in a large number of trials.

In a contrasting approach to the one taken by [49], Le Cam [99] in 1953 gave conditions for when a Bayes estimator is consistent for suitable prior distributions. Le Cam proved that in certain cases consistency of the maximum likelihood estimator implies consistency of the Bayes estimator. Some partial generalisations of this result are provided by Bickel & Yahav [14], who show the result applies

to a richer class of loss functions. Strasser [128] takes this further by proving that every possible set of conditions which implies consistency of the maximum likelihood method also implies consistency of Bayes estimates for a large class of priors. As before, however, all results are confined to the case of i.i.d. observations. Moreover, the result still necessitates proving consistency of the maximum likelihood estimate first.

Another notable contribution comes from Barron et al. [7], who prove consistency by showing that, for all values of ε , the probability of the posterior density lying within an ε -neighbourhood of the true density tends to 1 as more data is observed. They use the Hellinger distance to define the ε -neighbourhood and use a finite-dimensional set of densities to approximate the true set of posterior densities. Other authors, such as Shen & Wasserman [124] and Ghosal et al. [73], have taken the analysis further by investigating the rates of convergence of the posterior distributions around the true parameter. However, again, results are limited to the case of i.i.d. data.

2.2.2 Single Price Observation

With reference to (2.2), consider observing only one option price $f_t^{(1)}(\theta)$ at each observation time t , i.e. let $I_t = I = \{1\}$ in the notation of Section 2.1.1. Denote by θ^* the true value of the parameter. Assume that the market noises are given by

$$e_t^{(1)} \sim N(0, (\varepsilon_t^{(1)})^2)$$

for $\varepsilon_t^{(1)} > 0$ and are independent of each other but not necessarily identically distributed. Since we are only considering one option we drop the superscripts. Take a nested sequence of partitions $\Upsilon_n \supset \Upsilon_{n-1}$, where recall

$$\Upsilon_n = \Upsilon_n([0, T]) = \{t_1, \dots, t_n : 0 = t_1 < t_2 < \dots < t_n \leq T\},$$

and define

$$\mathcal{G}_{t_n} = \sigma(\{e_s : s \in \Upsilon_n\}),$$

i.e. the sigma-field generated by the noise random variables until time t_n . So in particular $\mathcal{G}_{t_n} \supset \mathcal{G}_{t_m}$ for all $n \geq m$.

Assumption 2.2.3. $\mathcal{F}_{t_n} \perp \mathcal{G}_{t_m}$ for all (n, m) , i.e. the driving process of the underlying is independent from the market noise. This is a reasonable assumption to make since observational errors are unlikely to be correlated to fundamental economic phenomena.

Let $V = \{V_t : t \in \Upsilon_\infty\}$ with V_t given by (2.2), and write $\theta_n(V) := \theta|\mathcal{F}_{t_n} \vee \mathcal{G}_{t_n}$ for shorthand. $\theta_n(V)$ is the random variable corresponding to the unknown parameter θ , conditioned on the price

observations up to time t_n . In other words, $p(\theta_n(V))$ is the posterior density of θ after n observations and is explicitly given by

$$\begin{aligned} p(\theta_n(V)) = p_n(\theta|V) &= \frac{p_n(V|\theta)p(\theta)}{p_n(V)} \\ &= \frac{p(V_{t_1}|\theta)\dots p(V_{t_n}|\theta)p(\theta)}{p_n(V)} \\ &= \prod_{t \in \Upsilon_n} \frac{1}{\sqrt{2\pi\varepsilon_t}} \exp\left\{-\frac{1}{2\varepsilon_t^2}(V_t - f_t(\theta))^2\right\} \frac{p(\theta)}{p_n(V)}. \end{aligned}$$

Define the sequence of Bayes estimators $\hat{\theta}$ by

$$\begin{aligned} g(\theta_n(V), \theta') &= \mathbb{E}[L(\theta_n(V), \theta')] \\ &= \int_{\theta} L(\theta, \theta') p_n(\theta|V) d\theta \\ \hat{\theta}_n(V) &= \arg \min_{\theta' \in \Theta} \{g(\theta_n(V), \theta')\}, \end{aligned}$$

where $L(\theta, \theta')$ is a loss function, Θ is the support for the prior density $p(\theta)$ of θ and $p_n(V)$ is described by (2.7). Note that $\hat{\theta}_n(V)$ is not necessarily unique.

We are interested in the distribution of $\theta_n(V)$, and in particular $\hat{\theta}_n(V)$. Note that the distribution of $\theta_n(V)$ could be numerically found by interpreting (2.2) as a discrete nonlinear particle filtering problem [70] with not necessarily identical f_t and not necessarily identically distributed noises e_t . However, for the purposes of our approach it is informative to prove the consistency result from first principles.

Assumption 2.2.4. The prior $p(\theta)$ and its support Θ satisfy:

- i) $p(\theta)$ is continuous at θ^* ,
- ii) $p(\theta^*) > 0$,
- iii) $p(\theta)$ is finite on Θ ,
- iv) Θ is closed and bounded.

Assumption 2.2.5. \exists finite $k > 0$ such that, for all $\theta, \theta' \in \Theta$, $|\theta - \theta'| > 0$, $\forall t$, the functions f_t and noise standard deviations ε_t , conditional on \mathcal{F}_t and \mathcal{G}_t respectively, satisfy

$$\frac{1}{\varepsilon_t} \frac{|f_t(\theta) - f_t(\theta')|}{|\theta - \theta'|} \geq k.$$

Note the fourth condition of Assumption 2.2.4 implies the support Θ is compact because Θ is finite-dimensional. Assumption 2.2.5 implies all the f_t are strictly monotone, which is a reasonable assumption to make. We also expect the noise standard deviation to be finite. We see later in Example 2.2.8 that these assumptions are satisfied in the Black-Scholes model for European call price functions.

Lemma 2.2.6. $\theta_n(V) \xrightarrow{P} \theta^*$.

Proof. Fix f_t for all $t \in \Upsilon_\infty$, so that ‘ P ’ refers only to the randomness caused by the market noise since we have fixed the stock price process up to time t_n (by fixing f_t). Then, by the definition of the posterior,

$$p_n(\theta|V) = q_n(V)e^{-\frac{1}{2}\phi_n(\theta,V)},$$

where $q_n(V)$ is a constant with respect to θ (normalising the posterior density) and

$$\begin{aligned} \phi_n(\theta, V) &= -2 \log p(\theta) + \sum_{t \in \Upsilon_n} \frac{1}{\varepsilon_t^2} (V_t - f_t(\theta))^2 \\ &= -2 \log p(\theta) + \sum_{t \in \Upsilon_n} \left(\frac{V_t - f_t(\theta^*)}{\varepsilon_t} - \frac{f_t(\theta) - f_t(\theta^*)}{\varepsilon_t} \right)^2 \\ &= -2 \log p(\theta) + \sum_{t \in \Upsilon_n} \left(Z_t - \frac{f_t(\theta) - f_t(\theta^*)}{\varepsilon_t} \right)^2 \end{aligned}$$

for all θ such that $p(\theta) > 0$, and where the Z_t are independent standard Gaussian random variables. By the first and second properties of Assumption 2.2.4, there exists $\epsilon_0 > 0$ such that $p(\theta) \geq d > 0$ for all $\theta \in [\theta^* - \epsilon_0, \theta^* + \epsilon_0]$. By the fourth property of Assumption 2.2.4, $\Theta \subset [\underline{\theta}, \bar{\theta}]$ for some constants $\underline{\theta}, \bar{\theta} \in \mathbb{R}$. Take $[\underline{\theta}, \bar{\theta}]$ as the support for the posterior by extending by zero the posterior measure from Θ to the interval $[\underline{\theta}, \bar{\theta}]$.

Now for all $\epsilon > 0$ such that $\epsilon_0 > \epsilon$,

$$\int_{\underline{\theta}}^{\theta^* - \epsilon} p_n(\theta|V) d\theta = \int_{\underline{\theta}}^{\theta^* - \epsilon} q_n(V)e^{-\frac{1}{2}\phi_n(\theta,V)} d\theta \leq [(\theta^* - \epsilon) - \underline{\theta}] q_n(V)e^{-\frac{1}{2}\phi_n(\alpha_n, V)},$$

where

$$\alpha_n = \arg \max_{[\underline{\theta}, \theta^* - \epsilon]} e^{-\frac{1}{2}\phi_n(\theta, V)}.$$

Also observe that

$$\begin{aligned} q_n(V) &= \frac{1}{\int_{\Theta} e^{-\frac{1}{2}\phi_n(\theta, V)} d\theta} \\ &\leq \frac{1}{\int_{\theta^*}^{\theta^* - \epsilon/2} e^{-\frac{1}{2}\phi_n(\theta, V)} d\theta} \\ &\leq \frac{2}{\epsilon} e^{\frac{1}{2}\phi_n(\beta_n, V)}, \end{aligned}$$

where

$$\beta_n = \arg \min_{[\theta^* - \epsilon/2, \theta^*]} e^{-\frac{1}{2}\phi_n(\theta, V)}.$$

Combining the two inequalities gives

$$\int_{\underline{\theta}}^{\theta^* - \epsilon} p_n(\theta|V) d\theta \leq \frac{[(\theta^* - \epsilon) - \underline{\theta}]}{\epsilon} e^{-\frac{1}{2}[\phi_n(\alpha_n, V) - \phi_n(\beta_n, V)]},$$

where

$$\begin{aligned} & \phi_n(\alpha_n, V) - \phi_n(\beta_n, V) \\ &= -2 \log p(\alpha_n) + 2 \log p(\beta_n) + \sum_{t \in \Upsilon_n} \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\theta^*)}{\varepsilon_t} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\theta^*)}{\varepsilon_t} \right)^2 \right]. \end{aligned} \quad (2.8)$$

But by the assumptions we have shown that $p(\beta_n) \geq d > 0$ and $p(\alpha_n)$ is finite using the third property of Assumption 2.2.4. Hence we can bound the first part of (2.8),

$$-2 \log p(\alpha_n) + 2 \log p(\beta_n) > -\infty.$$

Expanding the second part of (2.8) and then using $\alpha_n < \beta_n \leq \theta^*$ and f_t is strictly monotone gives

$$\begin{aligned} & -2 \sum_{t \in \Upsilon_n} Z_t \left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{\varepsilon_t} \right) + \sum_{t \in \Upsilon_n} \left[\left(\frac{f_t(\alpha_n) - f_t(\theta^*)}{\varepsilon_t} \right)^2 - \left(\frac{f_t(\beta_n) - f_t(\theta^*)}{\varepsilon_t} \right)^2 \right] \\ & \geq -2 \sum_{t \in \Upsilon_n} Z_t \left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{\varepsilon_t} \right) + \sum_{t \in \Upsilon_n} \left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{\varepsilon_t} \right)^2 \\ &= X_n, \end{aligned} \quad (2.9)$$

where $X_n \sim N(s_n, 4s_n)$ is a normal random variable with

$$s_n = \sum_{t \in \Upsilon_n} \left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{\varepsilon_t} \right)^2.$$

Now note that, by the definitions of α_n and β_n , we must have $|\alpha_n - \beta_n| \geq \epsilon/2$. Assumption 2.2.5 thus implies

$$s_n \geq \sum_{t \in \Upsilon_n} k^2 |\alpha_n - \beta_n|^2 \geq nk^2 \frac{\epsilon^2}{4} \rightarrow +\infty.$$

So observe that, for any $r \in \mathbb{R}^+$, the probability

$$\begin{aligned} \mathbb{P}[e^{-\frac{1}{2}X_n} \geq r] &= \mathbb{P}[X_n \leq -2 \log r] = \int_{-\infty}^{-2 \log r} \frac{1}{\sqrt{8\pi s_n}} e^{-\frac{1}{8s_n}(x-s_n)^2} dx \\ &= \Phi\left(\frac{-2 \log r - s_n}{\sqrt{4s_n}}\right) \\ &\rightarrow \Phi(-\infty) \\ &= 0 \end{aligned}$$

as $n \rightarrow \infty$, where Φ is the cumulative distribution function for a standard normal random variable.

Thus, we have

$$\int_{\underline{\theta}}^{\theta^* - \epsilon} p_n(\theta | V) d\theta \rightarrow 0.$$

and by an identical line of argument we can show that

$$\int_{\theta^* + \epsilon}^{\bar{\theta}} p_n(\theta | V) d\theta \rightarrow 0$$

for all $\epsilon > 0$ such that $\epsilon_0 > \epsilon$. But since $p_n(\theta|V)$ is a probability density this implies that, for all $\epsilon > 0$ such that $\epsilon_0 > \epsilon$,

$$\int_{\theta^* - \epsilon}^{\theta^* + \epsilon} p_n(\theta|V) d\theta \rightarrow 1$$

as $n \rightarrow \infty$. So we conclude that, in the limit as $n \rightarrow \infty$, the posterior probability measure converges to an atomic measure $\mathbb{P}_{\delta_{\theta^*}}$ such that $\mathbb{P}_{\delta_{\theta^*}}[\{\theta^*\}] = 1$, i.e. a Dirac measure centered at θ^* . Hence, $\theta_n(V) \xrightarrow{P} \theta^*$. \square

Theorem 2.2.7. *For L bounded and continuous on Θ , the Bayes estimator $\hat{\theta}_n(V)$ is consistent.*

Proof. First observe that for a fixed θ we can write $L(\theta, \theta') = l_\theta(\theta - \theta')$ for some function $l_\theta : \mathbb{R} \rightarrow \mathbb{R}$. For $\delta > 0$

$$\begin{aligned} \mathbb{P}_{\theta^*}[|\hat{\theta}_n(V) - \theta^*| \geq \delta] &\leq \mathbb{P}_{\theta^*}[|\hat{\theta}_n(V) - \theta_n(V)| + |\theta_n(V) - \theta^*| \geq \delta] \\ &\leq \mathbb{P}_{\theta^*}[|\hat{\theta}_n(V) - \theta_n(V)| \geq \frac{1}{2}\delta] + \mathbb{P}_{\theta^*}[|\theta_n(V) - \theta^*| \geq \frac{1}{2}\delta], \end{aligned}$$

where \mathbb{P}_{θ^*} denotes probability with respect to the model with parameter θ^* . But $\mathbb{P}_{\theta^*}[|\theta_n(V) - \theta^*| \geq \frac{1}{2}\delta] \rightarrow 0$ as $n \rightarrow \infty$, since $\theta_n(V) \xrightarrow{P} \theta^*$ (Lemma 2.2.6), and

$$\begin{aligned} \mathbb{P}_{\theta^*}[|\hat{\theta}_n(V) - \theta_n(V)| \geq \frac{1}{2}\delta] &\leq \mathbb{P}_{\theta^*}[L(\hat{\theta}_n(V), \theta_n(V)) \geq \min\{l_{\hat{\theta}_n(V)}(\frac{1}{2}\delta), l_{\hat{\theta}_n(V)}(-\frac{1}{2}\delta)\}] \\ &\leq \frac{\mathbb{E}_{\theta^*}[L(\theta_n(V), \hat{\theta}_n(V))]}{\min\{l_{\hat{\theta}_n(V)}(\frac{1}{2}\delta), l_{\hat{\theta}_n(V)}(-\frac{1}{2}\delta)\}} \\ &\leq \frac{\mathbb{E}_{\theta^*}[L(\theta_n(V), \theta^*)]}{\min\{l_{\hat{\theta}_n(V)}(\frac{1}{2}\delta), l_{\hat{\theta}_n(V)}(-\frac{1}{2}\delta)\}} \\ &\rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$. In the second line we used Markov's Inequality [8], in the third line we used the minimality of $\hat{\theta}_n(V)$ and in the last line we used L is bounded and continuous with $\theta_n(V) \xrightarrow{D} \theta^*$. Hence, for all $\delta > 0$, $\mathbb{P}_{\theta^*}[|\hat{\theta}_n(V) - \theta^*| \geq \delta] \rightarrow 0$ as $n \rightarrow \infty$. \square

Example 2.2.8. Consider the Black-Scholes framework and associate θ with the constant Black-Scholes volatility parameter σ . Take the observed options to be European call prices so let $f_t(\sigma) = e^{-r(T-t)}\mathbb{E}[(S_T(\sigma) - K)^+ | \mathcal{F}_t]$. Then we know (e.g. from Bingham & Kiesel [16]) that f_t is differentiable everywhere on $[\sigma_{\min}, \sigma_{\max}]$ ($0 < \sigma_{\min}$ and $\sigma_{\max} < \infty$) and the derivative (vega) is given by

$$f'_t(\sigma) = S_t \sqrt{T-t} N'(\frac{\log(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma \sqrt{T-t}})$$

for non-dividend paying stock S where $N'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Note that for all combinations of S_t and K the vega is positive and bounded below by $f'_t(\sigma_{\min})$ on the interval $[\sigma_{\min}, \sigma_{\max}]$. Hence, assuming finite market noise, the Black-Scholes European call prices satisfy Assumption 2.2.5 and can be used as the observation option prices to get a consistent Bayes estimator for the Black-Scholes volatility (provided we also use a suitable prior satisfying Assumption 2.2.4).

2.2.3 Multiple Price Observations

Suppose now that $|I_t| > 1$ so we observe several option prices $f_t^{(i)}(\theta)$ for $i \in I_t$ at each observation time t . Again assume that market noises are independent (even those corresponding to the same observation time t) and normally distributed as before. Then the following corollary holds.

Corollary 2.2.9. (*To Theorem 2.2.7*). *For multiple price observation sets I_t and all bounded and continuous loss functions L , the Bayes estimator $\hat{\theta}_n(V)$ is consistent.*

Proof. The proof is identical to those for Lemma 2.2.6 and Theorem 2.2.7 except that the sigma-field generated by all the noise random variables up to time t_n is now

$$\mathcal{G}_{t_n} = \theta \left(\{e_s^{(i)} : s \in \Upsilon_n, i \in I_t\} \right),$$

all the summations are replaced as follows

$$\sum_{t \in \Upsilon_n} \longrightarrow \sum_{t \in \Upsilon_n, i \in I_t}$$

and values V_t, f_t, ε_t get superscripts to become $V_t^{(i)}, f_t^{(i)}, \varepsilon_t^{(i)}$ respectively. \square

Remark 2.2.10. The only real difference in the convergence of the estimator is that, for multiple observations, the convergence is likely to be faster since more information is added at each timestep to update the posterior $p_n(\theta|V)$.

2.2.4 Finite-Dimensional Non-Scalar Parameter

We generalise the previous results to the case when θ is not scalar but a finite-dimensional vector of parameters. We suppose that $\theta \in \mathbb{R}^M$ and assume a norm $\|\cdot\|$ on \mathbb{R}^M . The following assumption and theorem show that, with some modifications of the assumptions, it is not difficult to extend the previous results to the non-scalar case. We assume Assumption 2.2.4 as before, and replace Assumption 2.2.5 with the following.

Assumption 2.2.11. There exist finite $K > k > 0$ such that, for each $\theta \in \Theta$, $\exists n(\theta) \in \mathbb{N}$ such that

$$K^2 > \frac{1}{n} \sum_{t \in \Upsilon_n} \frac{1}{\varepsilon_t^2} \frac{|f_t(\theta) - f_t(\theta^*)|^2}{\|\theta - \theta^*\|^2} > k^2 \quad \forall n \geq n(\theta)$$

Note that Assumption 2.2.11 essentially states that for any $\theta \neq \theta^*$ the average observational error is eventually non-zero. For the scalar case, $n(\theta) = 1$ for all $\theta \in \Theta$.

Theorem 2.2.12. *For all L bounded and continuous on θ , the non-scalar Bayes estimator $\hat{\theta}_n(V)$ is consistent.*

Proof. The proof is similar to that of Lemma 2.2.6, with a few adjustments for the higher dimensionality. Instead of integrals on the real line we take integrals over M -dimensional space. So we consider

$$\int_{B(\theta^*, \epsilon)^c} p_n(\theta|V) d\theta$$

for all $\epsilon > 0$ and $\epsilon_0 > \epsilon$ where, for all θ such that $\|\theta - \theta^*\| \leq \epsilon_0$, $p(\theta) \geq d > 0$. $B(\theta^*, \epsilon)$ is the ball in \mathbb{R}^M with centre θ^* and radius ϵ and ‘ c ’ denotes the complement of a set in \mathbb{R}^M . Again we extend by zero the posterior measure for Θ to $\overline{\Theta}$. We get that the integral is bounded as

$$\int_{B(\theta^*, \epsilon)^c} p_n(\theta|V) d\theta \leq \frac{|\overline{\Theta}| - |B(\theta^*, \epsilon)|}{|B(\theta^*, \epsilon k/\sqrt{2K})|} e^{-\frac{1}{2}[\phi_n(\alpha_n, V) - \phi_n(\beta_n, V)]},$$

for the optimisers $\alpha_n, \beta_n \in \Theta$ defined similarly to before as

$$\alpha_n = \arg \max_{B(\theta^*, \epsilon)^c} e^{-\frac{1}{2}\phi_n(\theta, V)} \quad \beta_n = \arg \min_{B(\theta^*, \epsilon k/\sqrt{2K})} e^{-\frac{1}{2}\phi_n(\theta, V)}.$$

We then have an identical expression to (2.8) for $\phi_n(\alpha_n, V) - \phi_n(\beta_n, V)$. Again by Assumption 2.2.4, $p(\beta_n) > 0$ and $p(\alpha_n)$ is finite, so $-2 \log p(\alpha_n) + 2 \log p(\beta_n) > -\infty$. For the equivalent expression for (2.9), we use that, for $n \geq n(\alpha_n), n(\beta_n)$,

$$\begin{aligned} & \sum_{t \in \Upsilon_n} [(Z_t - \frac{f_t(\alpha_n) - f_t(\theta^*)}{\epsilon_t})^2 - (Z_t - \frac{f_t(\beta_n) - f_t(\theta^*)}{\epsilon_t})^2] \\ &= -2 \sum_{t \in \Upsilon_n} Z_t \frac{f_t(\alpha_n) - f_t(\beta_n)}{\epsilon_t} + \sum_{t \in \Upsilon_n} [(\frac{f_t(\alpha_n) - f_t(\theta^*)}{\epsilon_t})^2 - (\frac{f_t(\beta_n) - f_t(\theta^*)}{\epsilon_t})^2] \\ &\geq -2 \sum_{t \in \Upsilon_n} Z_t \frac{f_t(\alpha_n) - f_t(\beta_n)}{\epsilon_t} + nk^2 \epsilon^2 - nK^2(\epsilon k/\sqrt{2K})^2 \\ &= X_n \end{aligned}$$

where $X_n \sim N(\frac{1}{2}nk^2\epsilon^2, s_n)$ is a normal random variable with variance

$$\begin{aligned} s_n &= \sum_{t \in \Upsilon_n} (\frac{f_t(\alpha_n) - f_t(\beta_n)}{\epsilon_t})^2 = \sum_{t \in \Upsilon_n} (\frac{f_t(\alpha_n) - f_t(\theta^*)}{\epsilon_t} + \frac{f_t(\theta^*) - f_t(\beta_n)}{\epsilon_t})^2 \\ &\leq 3 \sum_{t \in \Upsilon_n} (\frac{f_t(\alpha_n) - f_t(\theta^*)}{\epsilon_t})^2 + 3 \sum_{t \in \Upsilon_n} (\frac{f_t(\theta^*) - f_t(\beta_n)}{\epsilon_t})^2 \\ &\leq 3(nk^2\epsilon^2 + \frac{1}{2}nk^2\epsilon^2). \end{aligned}$$

Hence, as before we can show that for any $r \in \mathbb{R}^+$, $\mathbb{P}[e^{-\frac{1}{2}X_n} \leq r] \rightarrow 0$.

All other arguments follow as for the rest of the proof of Lemma 2.2.6 and Theorem 2.2.7 \square

The above result can easily be generalised to the case of multiple observations as before.

2.3 Local Volatility Model Formulation

In this section we introduce the local volatility model and its associated calibration problem. Although there is debate surrounding whether to calibrate volatility according to historical data or

to current known market data (see for example [116] for a detailed discussion), common practice is usually the latter. In this chapter we adhere to the common market practice, but explain the difficulties endemic to this calibration approach.

Consider the model originally proposed by Black & Scholes [17] with finite time horizon T . Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$ be a filtered probability space and $(Z_t)_{0 \leq t \leq T}$ a Brownian motion such that \mathcal{F}_t is the natural filtration of Z_t over Ω and $\mathcal{F} = \mathcal{F}_T$. Then the underlying asset price S is assumed to follow geometric Brownian motion, i.e. satisfies the stochastic differential equation (SDE) under the physical measure \mathbb{P} :

$$dS_t = \mu S_t dt + \sigma S_t dZ_t, \quad (2.10)$$

where the drift μ is the expected growth rate and the diffusion coefficient σ is the volatility. The original Black-Scholes model takes μ and σ as constants, but the model has subsequently been extended so that they can also be deterministic functions of time, and σ can even be a deterministic function of both time and asset price. Standard no-arbitrage arguments (see for example [85]) then show that the price V of an option written on S must satisfy the famous Black-Scholes partial differential equation (PDE)

$$\frac{\partial V}{\partial t} + rS \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV = 0, \quad (2.11)$$

where r is the rate of interest and we assume zero dividends. [17] analytically solves (2.11) to find the time-0 price for a variety of payoffs.

For example, the theoretical time-0 price of a non-dividend paying European call option with payoff $(S_T - K)^+$ is given by

$$EC(S_0, K, T, \sigma, r) = S_0 N(d_1) - K e^{-rT} N(d_2), \quad (2.12)$$

$$\text{where } d_1 = \frac{\log(\frac{S_0}{K}) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T}, \quad (2.13)$$

S_0 is the time-0 value of S and the volatility σ is constant (see for example [85] for a full derivation of (2.12)). In practice, we know the values for S_0, K, T, r , but cannot observe the volatility σ . We can however find the volatility implied by a given market price $V(K, T)$ for a European call option with strike K , maturity T and underlying with value S_0 at time 0. We denote this volatility as σ_{KT} , and call it the Black-Scholes *implied volatility*. Note that this relationship is one-to-one, that is, for any call price $V(K, T)$ there exists one and only one implied volatility σ_{KT} . In fact, this relationship holds for all convex payoffs.

The Black & Scholes model assumes volatility to be constant for all options. However, for a set of market European call prices for example, the implied volatilities are found to vary with both strike and maturity (see for example [116] or [110]). We call the variation with respect to strike the *skew*

or *smile* and the variation with respect to maturity the *term structure*. To model these effects, one obvious choice for σ is

$$\sigma = \sigma(S, t), \quad (2.14)$$

i.e. a deterministic function that varies with both asset price S and time t . We call this choice for σ the *local volatility* function. It is called ‘local’ because it assigns an instantaneous volatility to each state (S, t) . The variability for different times and asset prices allows one to capture the smile and term structure aforementioned. The question remains however of what this function looks like and how to find the form given by (2.14).

2.3.1 Local Volatility Assumptions

Before looking at how we might recover the implied local volatility function, we should discuss what we would expect the local volatility function to look like. There are three main properties that we would expect the local volatility surface to exhibit:

Positivity $\sigma(S, t) > 0$ for all values of S and t ; since the squared price variation $\sigma^2 > 0$ we adopt the convention $\sigma > 0$.

Smoothness there should be no sharp spikes or troughs in the surface; as Jackson et al. [89] point out, there is no reason for expecting current prices to be capable of accurately predicting abrupt changes in future volatility.

Consistency for small values of t especially, σ should be close to today’s at-the-money volatility σ_{atm} ; at-the-money volatility is usually equated with the Black-Scholes implied volatility for a European call option with strike equal to the current price of the underlying and for a short maturity. As a consequence of the smoothness property, today’s at-the-money volatility will roughly determine the position of the local volatility surface in \mathbb{R}^3 .

2.3.2 Dupire’s Formula

In 1994 Dupire [51] showed that, if we can observe the market prices (equivalently implied volatilities) for European call options for all strikes and maturities, then the local volatility (2.14) can be uniquely specified. For $V(K, T)$ the market price of a European call with strike K and maturity T , Dupire showed that the calibrated local volatility is given by

$$\sigma(S, t) = \sqrt{2 \frac{\frac{\partial V}{\partial T}(S, t) + rS \frac{\partial V}{\partial K}(S, t)}{S^2 \frac{\partial^2 V}{\partial K^2}(S, t)}}, \quad (2.15)$$

assuming $V(K, T)$ is twice differentiable with respect to K and once differentiable with respect to T , the second derivative with respect to K is nowhere zero and the radicand is non-negative. A full derivation of this result is given in Appendix C.1.

Although Dupire's formula is a neat theoretical result, in practice it is not very easy to use. The problem is that in reality we do not have the market prices for all strikes and maturities. Hence, the assumptions made about the existence of the partial derivatives in (2.15) are invalid. So one way to implement Dupire's formula would be to first interpolate market prices to find the full function $V(K, T)$ and make it appropriately differentiable. However, there is no obvious way to interpolate the prices in such a way that the radicand in (2.15) remains positive and finite. And, as Berestycki et al. [10] comment, for short maturities especially, the resulting surface is too sensitive to changes in the choice of interpolant and hence the method is not robust.

Furthermore, attempts made to directly implement Dupire's formula for a given set of market prices have recovered local volatility functions that are non-smooth and exhibit large spikes. For example, the plots in Crepey's paper [38] of the squared local volatility surface obtained by application of Dupire's formula (2.15) to DAX data show surfaces with highly irregular and unstable behaviour. Rebonato's book [116] also has plots of local volatility surfaces calculated using Dupire's formula and they exhibit the same spiked and unstable features. As discussed, there is no reason why today's prices should accurately forecast abrupt changes in future volatility.

The conclusion is that application of the Dupire formula is impractical and gives results that are often unusable for pricing and hedging purposes.

2.3.3 Error Functional Minimisation Approach

Since we do not have enough (market) data to analytically compute the local volatility function, we must instead use numerical methods to find the best solution we can from the data we do have. Using the notation from earlier in this chapter, for $i \in I$, let $V^{(i)}$ be the market price for a European call option with strike K_i , maturity T_i , on underlying S with value S_0 at time 0. We could calibrate the local volatility to any market prices we like, but most widely used in practice and literature are European calls because of their simplicity and availability. As remarked at the beginning of this chapter, the market does not usually give a single price $V^{(i)}$, but only defines it to within a bid-ask spread $[V^{(i)bid}, V^{(i)ask}]$. For purposes of calibration it is usual (e.g. [89]) to then assume the fair market price is

$$V^{(i)} = \frac{1}{2}(V^{(i)bid} + V^{(i)ask}). \quad (2.16)$$

By calibrating $\sigma(S, t)$ we specifically want to find $\sigma : \mathbb{R}^+ \times [0, T] \rightarrow \mathbb{R}^+$ which minimises an error functional, such as the one proposed by Jackson et al. [89],

$$G_1(\sigma) = \sum_{i \in I} w_i |f^{(i)}(\sigma) - V^{(i)}|^2, \quad (2.17)$$

where $f^{(i)}(\sigma) = EC(S_0, K_i, T_i, \sigma, r)$ and by abuse of notation this is the theoretical value for a European call option priced in the local volatility model rather than in the constant volatility Black-Scholes model. Convention is to have the weights $w = \{w_i : i \in I\}$ satisfy $\sum_{i \in I} w_i = 1$. Weighting

different error terms is useful to give priority in calibration to those options that are more heavily traded or have greater liquidity [89].

The calibration formulation is simple and intuitive. The formulation given by (2.17) seeks to find the function σ which minimises the difference between the theoretical and observed prices. However, with reference to the theory given on inverse problems in Section 2.1, this minimisation problem is ill-posed.

2.3.4 Ill-Posedness of Local Volatility Calibration

Recall from Section 2.1.2 Hadamard's criteria i), ii), iii) for well-posedness. Since we are trying to recover a function σ , the inverse problem is automatically infinite-dimensional. Hence, given a finite number of arbitrage-free prices, a solution will usually always exist (for example by using Dupire's formula (2.15)) and we do not violate i). However, it is usually the case that more than one surface σ will be able to recover the set of observed prices, which violates ii). Moreover, using only a finite dataset of prices, a solution is unlikely to be stable with respect to changes in the observed prices, which violates iii). To see this consider Dupire's formula (2.15) as an example. Firstly the formula requires computing derivatives from discrete observed data, which is unstable, and secondly the denominator in the radicand can become very small for far in and out of the money European call options which again makes the solution unstable. Hence, either some sort of regularisation technique described in Section 2.1.3 is needed to make the minimisation problem convex or a different method is needed to specify a unique and continuous solution.

2.3.5 Literature on Local Volatility Calibration

In this section we review the literature written on the subject of local volatility calibration. As we shall see, the most common way of solving the calibration problem is through regularisation techniques. However, there is also a variety of alternative fixes that have been proposed in the academic literature and merit consideration.

In Section 2.1.3 we saw how an ill-posed inverse problem can be recast as a minimisation problem with an added regularisation term to induce convexity. This was desirable to guarantee uniqueness and stability of the solution. In the papers we review, the stabilising term added to the error functional (2.17) is sometimes referred to as the *cost functional* or *penalty functional* or *smoothness functional* but they are all equivalent. For what follows, we take $G_1(\sigma)$ as defined by (2.17).

Lagnado & Osher [98] choose the square of the L_2 norm of the gradient of σ as the regularisation term. Hence in their paper they minimise the functional

$$F_{lagosh}(\sigma) = G_1(\sigma) + \lambda \|\nabla\sigma\|_2^2, \quad (2.18)$$

where the regularisation parameter λ is a constant (chosen by trial and error to optimise the rate of convergence of their numerical minimisation procedure) and $w_i = 1/|I| \forall i \in I$. By taking $\|\nabla\sigma\|_2^2$ as

the regularisation term, they find σ that is smooth and close to minimising $G_1(\sigma)$. The authors use a gradient-descent scheme to find the optimal σ , testing $Flagosh(\sigma)$ for each iteration of σ generated by the scheme. They use a finite-difference method for solving the associated Black-Scholes PDE (2.11) with appropriate boundary condition (payoff). Although convergence of their procedure to the optimal solution of (2.18) is not proved, their numerical results for simulated data and market data taken from S&P 500 show that good calibration is achieved, especially as the number of price observations is increased. However, Lagnado & Osher's method has some drawbacks. On the numerics side, the minimisation requires calculating variational derivatives using a form of (2.11) which is computationally expensive. On the financial side, the local volatility surface $\sigma(S, t)$ is only generated for several discrete time points for values of S close to the money, and thus is difficult to use for pricing more complex options. Furthermore, the uniformity of the weights w_i does not account for the varying importance of options more heavily traded or more liquid.

Chiarella et al. [26] try to reduce the computation complexity of Lagnado & Osher's method by using fewer references to variational derivatives and making approximations using the Black-Scholes formula (2.12). Their method requires fewer iterations and is thus computationally faster. However, [26] still produces volatility surfaces which cannot be used to price exotic options that depend on far out-of-the money values of volatility, such as barrier options.

Jackson et al. [89] present a more direct regularisation approach that avoids the need for computing variational derivatives. [89] first represents the local volatility surface by a set of nodes, which are later interpolated using natural cubic splines. Weights are chosen to reflect the priority in calibration so, in particular, at-the-money options are given much greater weight. The method is then again to minimise the functional $Flagosh$, albeit in the guise of a discretised version of the $\|\nabla\sigma\|_2^2$ term. A quasi-Newton algorithm is used for optimisation, and a piecewise quadratic finite-element method is used for solving the Black-Scholes PDE at each iteration. These techniques reduce minimisation time on a computer dramatically. However, there are some disadvantages of their approach. Firstly, the regularisation parameter λ is still arbitrarily chosen. Although selected to maximise convergence, the choice of λ lacks any financial interpretation. Secondly, the method is only demonstrated for a relatively low nodes-to-prices ratio - 15 nodes are calibrated to only 10 prices. In reality we would expect to calibrate to between 50 and 100 prices, and this is unlikely to be as easily done in the method presented. For example, for 100 prices using 150 nodes would make the quasi-Newton algorithm run very slowly. Thirdly, their method is susceptible to over-regularisation, as they seek to show uniqueness of the solution. This is why, for example, we see that two of the pricing errors are relatively high — between 4 and 7 basis points.

Work has also been carried out on the more theoretical side, looking at stability and rates of convergence for methods trying to minimise the functionals of type (2.18). Crepey [37] considers

Tikhonov regularisation and proves stability and convergence results. His approach is to specify a so-called prior (though not strictly in the Bayesian probabilistic sense) σ_0 and employ a regularisation term that minimises the difference between this and the calibrated σ . [37]’s proposed minimisation functional is then

$$F_{crepey}(\sigma) = G_1(\sigma) + \lambda (\|\sigma - \sigma_0\|_2^2 + \|\nabla\sigma\|_2^2). \quad (2.19)$$

Note that the norm of both σ and its derivative $\nabla\sigma$ appears in the regularisation term. More recently, Egger & Engl [52] have followed a similar route, coupling Tikhonov regularisation with a prior guess for σ , and using the same functional F_{crepey} to minimise.

Although regularisation by an appropriate (Tikhonov) smoothing functional is the classical way to solve the ill-posed inverse problem of calibration in the applied mathematics literature, it is by no means the only one. In the remainder of this section we survey some other techniques for calibrating the diffusion process to market prices, and highlight some of their advantages and shortcomings.

Recall the definition for the Black-Scholes implied volatility σ_{KT} of a call option given at the beginning of Section 2.3. Given a set of call options with varying maturity and strike, we can easily extract the implied volatility surface

$$\sigma_{imp}(K, T) = \sigma_{KT}, \quad (2.20)$$

albeit through some extra interpolation that can be problematic in the sense that the resulting surface does not give arbitrage-free prices. We have already reviewed in Section 2.3.2 the relationship found by Dupire between the implied volatility surface and local volatility surface. Rebonato [116] and Derman et al. [46] offer some qualitative conclusions and rules of thumb. Carr & Madan [24] use a volatility smile for a fixed maturity to find a pricing function and then invert this to recover the local volatility. Berestycki et al. [10] first specify that the local volatility $\sigma \in \text{BUC}$, where BUC is the space of bounded and uniformly continuous functions, and then find a PDE linking σ with σ_{imp} , in a similar fashion to the Dupire PDE linking σ with European call prices. The requirement that $\sigma \in \text{BUC}$ and an asymptotic formula for σ_{imp} in terms of σ near expiry regularises the problem. In both methods, however, implicit is the assumption that we have a continuum of implied volatilities which in reality we do not.

Another popular method for recovering local volatility is to use tree and lattice algorithms (see [116] for a full treatment). The basic idea is to construct a tree, the nodes of which represent possible values attainable by an asset price S at different times t , and recursively calculate values of the volatility at these nodes using market data. Dupire [51], Rubinstein [118] and Dumas et al. [50] use binomial trees, while Derman et al. [45], Crepey [38] and Britten-Jones & Neuberger [21] employ trinomial trees. Trinomial trees allow greater flexibility in deciding where to position nodes which is advantageous in matching trees to smiles. Although the tree algorithms described by these papers are fast and efficient, the results can be unsatisfactory. Rebonato [116] shows the Derman & Kani

method can lead to spiked local volatility surfaces and non-flattening wings, neither of which are expected or desirable properties of the local volatility surface. Jackson et al. [89] also point out that such tree methods only find instantaneous volatility for a triangular section of space which inhibits accurately pricing some exotics. There is also again no guarantee that resulting prices are arbitrage-free. Moreover, with the exception of Crepey’s method, none of the tree algorithms address the fundamental calibration problem of ill-posedness.

Avellaneda et al. [4] specify a Bayesian prior for σ and then use relative-entropy minimisation to find the local volatility surface which reproduces observed market prices and is closest, in terms of the Kullback-Leibler information distance, to this prior. Bodurtha & Jermakyan [18] consider this work with a small-parameter power expansion of the local volatility function and numerically solve the Black-Scholes PDE with market prices to find the coefficient functions in the power expansion. Although the methodology is interesting and useful, the numerical results are unconvincing: in both papers, the surfaces produced exhibit the spikes and troughs which we view as unrealistic.

Egger et al. [53] try to simplify the calibration problem by decoupling the smile and term structure, so that the local volatility is expressible as

$$\sigma(S, t) = \sigma_1(S) \sigma_2(t). \quad (2.21)$$

Unfortunately, it is unlikely this approach will be useful as a general framework for finding the local volatility function, since market data show that the shape of the volatility smile is not consistent over time. For example, for options on equity prices the smile tends to flatten as maturity increases [116].

Coleman et al. [28] approximate the local volatility surface by a bicubic spline and match the position (S, t) of the spline knots to the strike-maturity coordinates (K, T) of the market prices. They put bounds on the values of the volatility at each knot to restrict the space of solutions. However, they use a large ratio of knots to prices, so the computational cost is considerable — for example, 70 spline knots are used in order to calibrate to 70 market prices.

A recent article by Hamida & Cont [79] is closer to the work we seek to do in this thesis. In their paper Hamida & Cont first specify the properties they would like of σ , smoothness and positivity, and represent these via a prior Gaussian probability density; so a smoother σ has greater density. They then draw samples from this prior density for σ and use evolutionary optimisation to select those σ which reproduce market prices to within a chosen error tolerance level δ , where δ is chosen as a weighted average of the bid-ask spreads of the prices. In this way they find many calibrated local volatility surfaces, some with striking differences from others. They finish by making conclusions about the implied model uncertainty, measured using the framework set out in earlier work by Cont [30]. One criticism is that their method does not take full advantage of all the information available. The evolutionary optimisation does not provide any way of comparing the different solutions, but only gives a set of surfaces which reproduce market prices to within δ .

Hence, posterior distributions cannot be calculated. And by not explicitly sampling the posterior distribution, [79] cannot evaluate confidence intervals for prices and optimal hedges. By recasting the problem in a formal Bayesian framework, this is one of the gaps this thesis hopes to fill.

2.4 Bayesian Modelling

In this section we propose concrete expressions for the Bayesian prior, likelihood and posterior functions defined in Section 2.1.4. We then apply these expressions to the local volatility model. Again, for what follows we assume the model parameter θ is finite-dimensional, so the working example is a discretised local volatility surface.

2.4.1 Prior Distribution (Regularisation)

For the purposes of introducing the theory we consider the simplest density — the Gaussian density. It is also the locally second-order approximation to most smooth densities. The approach is to reformulate prior beliefs of θ into a norm (*cost function*) $\|\cdot\|$ of θ so that parameters which better satisfy the prior beliefs have smaller norm. The natural Gaussian prior is

$$p(\theta) \propto \exp \left\{ -\frac{1}{2} \tilde{\lambda} \|\theta - \theta_0\|^2 \right\}, \quad (2.22)$$

where θ_0 is some initial value of θ and $\tilde{\lambda}$ is a constant which quantifies how strong our prior assumptions are: a higher value of $\tilde{\lambda}$ indicating greater confidence in our assumptions. $1/\tilde{\lambda}$ can be thought of as the prior variance of θ . The choice of confidence parameter $\tilde{\lambda}$ is discussed in Section 2.4.4. From (2.22) we see that those θ which better satisfy prior beliefs have greater density. To clarify the choice of norm $\|\cdot\|$ (which entirely captures our prior beliefs), we continue our example of the local volatility model.

In light of the assumptions for the local volatility model presented in Section 2.3.1, we take for our prior

$$p(\sigma) \propto \exp \left\{ -\frac{1}{2} \tilde{\lambda} \|\log(\sigma) - \log(\sigma_{atm})\|_\kappa^2 \right\}, \quad (2.23)$$

where $\|\cdot\|_\kappa$ is a norm. Working in the logarithmic space guarantees σ is positive and the norm ensures greater prior density is attached to σ that are both smoother and closer to ATM volatility σ_{atm} .

The type of smoothness will depend on how we choose $\|\cdot\|_\kappa$. Following the regularisation functional used, for example by Fitzpatrick [62], we choose the following variation of the Sobolev norm $\|\cdot\|_{1,2}$:

$$\|u\|_\kappa^2 = (1 - \kappa) \|u\|_2^2 + \kappa \|\nabla u\|_2^2, \quad (2.24)$$

where $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)$ is the grad operator, $\|\cdot\|_2$ is the standard L_2 norm and $\kappa \in (0, 1)$ is a pre-specified constant.

2.4.2 Likelihood Function (Calibration Error)

Recall $V_t^{(i)}$ is the market observed price at time t of a European call with strike K_i and maturity T_i written on underlying S taking value S_t at time t , and $f_t^{(i)}(\theta)$ is the theoretical price for the same derivative when the model parameter is θ . Using the terminology of Jackson et al. [89], we define the basis point square-error function as

$$G_t(\theta) = \frac{10^8}{S_t^2} \sum_{i \in I} w_i |f_t^{(i)}(\theta) - V_t^{(i)}|^2,$$

where the w_i are pre-specified weights summing to one. Note the smaller the value of $G(\theta)$ the more likely the value of θ .

Since $V_t^{(i)}$ is usually only observed to within its bid-ask spread, $[V_t^{(i)bid}, V_t^{(i)ask}]$, define $\delta_i = \frac{10^4}{S_0} |V_t^{(i)ask} - V_t^{(i)bid}|$ as the basis point bid-ask spread for the i^{th} option. Then we shall only attach positive Bayesian posterior weight to parameters θ which on average reproduce prices to within the average basis-point bid-ask spread. In other words, we will attach positive likelihood to θ only if

$$G(\theta) \leq \delta^2 \tag{2.25}$$

where $\delta^2 = \sum_{i \in I} w_i \delta_i^2$ is the pre-specified average basis point square-error tolerance. Note that we could modify the likelihood so that every price is calibrated to within its tolerance, i.e. $(10^8/S_t^2) |f_t^{(i)}(\theta) - V_t^{(i)}| \leq \delta_i^2$ for all $i \in I$, but fitting every tolerance would be computationally much more difficult. Essentially we are modelling the basis-point error for the i^{th} option as

$$\frac{10^4}{S_0} (f_t^{(i)}(\theta^*) - V_t^{(i)}) \sim N(0, \delta_i^2), \tag{2.26}$$

as we did in Section 2.2 where θ^* is the true parameter value. And for the calculation of $G(\theta)$ we shall set

$$V_t^{(i)} = \frac{1}{2}(V_t^{(i)bid} + V_t^{(i)ask}).$$

Satisfaction of (2.25) is the approach also taken by Hamida & Cont [79]. However, they consider all surfaces σ satisfying the constraint (2.25) equally well calibrated, whereas in this paper we will differentiate between different degrees of calibration and sample the posterior distribution. For example, a surface θ which gives an average basis-point square-error of 1 is far better calibrated than one which gives an average basis-point square-error of 9.

Hence, for the Bayesian likelihood for non-parametric models we will take

$$p(V|\theta) \propto 1_{G(\theta) \leq \delta^2} \exp\left\{-\frac{1}{2\delta^2} G(\theta)\right\}. \tag{2.27}$$

So those surfaces σ which reproduce prices closest to the market observed prices V have the greatest likelihood values. Note that we have added δ^2 in the denominator of the fraction in the exponent to make the density look as Gaussian as possible. Observe from equation (2.26) that the basis point

errors are modelled as Gaussian with mean zero and variance δ_i , and that the weighted sum of variances is δ^2 . Hence we can think of $G(\theta)$ as truncated Gaussian with mean zero and variance δ^2 , so we write δ^2 in the numerator of the fraction.

2.4.3 Posterior Distribution

Combining the prior and likelihood functions we get the explicit form for the posterior function $p(\theta|V)$ as

$$p(\theta|V) \propto 1_{G(\theta) \leq \delta^2} \exp\left\{-\frac{1}{2\delta^2} [\lambda\|\theta - \theta_0\|^2 + G(\theta)]\right\}, \quad (2.28)$$

where $\lambda = \delta^2 \tilde{\lambda}$.

Remark 2.4.1. Observe that maximising the posterior (2.28) is equivalent to minimising the expression

$$\lambda\|\theta - \theta_0\|^2 + G(\theta), \quad (2.29)$$

over the set $\{\theta : G(\theta) \leq \delta^2\}$, and (2.29) is precisely the form of function authors such as Lagnado & Osher [98] and Jackson et al. [89] seek to minimise to find their optimal calibration parameter. This fact is also noted by Fitzpatrick [62]. This is not a coincidence but an insight into how the Bayesian approach reforms and generalises traditional Tikhonov regularisation methods into a unified framework.

2.4.4 Choice of Confidence Parameter

In this section we look at the role of λ in more detail and review some of the literature written on this parameter.

As noted in Remark 2.4.1, maximising the posterior (2.28) is equivalent to minimising the regularised error function (2.29). And the confidence parameter λ in (2.28) is called the *regularisation parameter* in the context of minimising regularised functions (see Section 2.1.3 for details). The literature that exists on optimal selection of the regularisation parameter is vast and dates back to the early 1970s. However, the choice of λ remains a very difficult question and is in general unsolved.

Authors such as Golub et al. [75] and Marquardt [103] looked at optimising the regularisation parameter in the linear regression model $y = X\beta + \epsilon$, where y is an n vector representing the observed data, X are underlying variables, ϵ is noise with mean zero. They sought to find a good value of the Lagrange multiplier λ for the problem of finding the minimum over β of

$$\frac{1}{n}\|y - X\beta\|^2 + \lambda\|\beta\|^2.$$

Hence, they refer to the regularised error function as a Lagrangian instead but the problem is the same. Various techniques such as minimum mean square error linear estimation, a priori estimation, minimax estimation have all been proposed for the linear regression model.

In 1984 Morozov published his book [109] on methods for solving ill-posed problems and introduced what is now widely referred to as the (*Morozov*) *Discrepancy Principle*. In the context of our problem the Discrepancy Principle essentially says that, if θ_λ is the minimiser of (2.29) and we restrict θ as being acceptable only if

$$G(\theta) \leq \delta^2, \quad (2.30)$$

then we should choose λ such that

$$G(\theta_\lambda) = \delta^2.$$

In words this says that we should choose λ as large as possible so that our solution θ_λ is still acceptable. Indeed, this is the method Cont & Tankov [31] take during their relative entropy approach to calibrating jump-diffusion models.

There are a number of theorems which show that the Discrepancy Principle gives a unique choice for λ depending on the properties of G (such as convexity) and δ^2 . However, this has not been the end of the story for the choice of regularisation parameter. For example, Engl [58] proposes some alternative a-posteriori Discrepancy Principle parameter choice methods for Tikhonov regularisation that give optimal convergence rates for linear ill-posed problems. By convergence rate, we specifically mean the rate by which the minimiser $\theta_\lambda \rightarrow \theta_0$ as $\delta^2 \rightarrow 0$, where $\lambda = \lambda(\delta)$. However, for the purposes of this thesis, expression (2.29) is nonlinear and we thus turn to other methods.

Engl et al. [60] improve the convergence rates of [58] for non-linear ill-posed problems and these rates are then optimised by Scherzer, Engl & Kunisch [121]. [121] uses a different a-posteriori regularisation parameter choice strategy which improves the convergence rate for restriction (2.30) to $\mathcal{O}(\delta^{4/3})$, compared with the rate of $\mathcal{O}(\delta)$ for the Morozov Discrepancy Principle. [121] finds that the Morozov Discrepancy Principle underestimates the value of the regularisation parameter. This shortcoming is further addressed by Scherzer [120]. Subsequent to the aforementioned results, which are exclusively for finite-dimensional ill-posed problems, papers by Weese [133] and Roths et al. [117] have extended the theory further to cover infinite-dimensional non-linear ill-posed problems.

As an alternative, Hansen [80] amongst others has advocated the use of L-curve analysis to determine the optimum regularisation parameter. Through experimentation, it has been found that a plot of the penalty function $\|\theta_\lambda\|$ versus the error function $G(\theta_\lambda)$ is often ‘L-shaped’ — exhibits a ‘corner’ in the bottom left of the graph above which the graph is very steep and to the right of which the graph is very flat. Advocates of L-curve analysis suggest using the value of λ at which this corner or kink occurs as a good value for the regularisation parameter. Intuitively, it is the value of λ for which the contribution from the penalty term and error term is balanced. However, so far very few formal treatments or satisfactory results using this approach have been found.

Although the choice of the regularisation parameter in Tikhonov regularised problems has been well studied, it is not immediately clear how to apply these results in the context of Bayesian posterior distributions. Recall that we are not trying to find the optimal value θ_λ but the optimal,

or rather, most representative, distribution $p(\theta|V)$. To this end, Fitzpatrick [62] has provided an appealing choice for λ .

Fitzpatrick first observes that the norm $\|\cdot\|^2$ in (2.29) induces a bilinear form on the parameter set Θ which gives rise to a matrix A such that

$$\|\theta - \theta_0\|^2 = (\theta - \theta_0)^T A^{-1}(\theta - \theta_0).$$

Hence we can write the posterior (2.28) as something that is closer to Gaussian:

$$p(\theta|V) \propto 1_{G(\theta) \leq \delta^2} \exp \left\{ -\frac{1}{2\delta^2} \left[\frac{\delta^2}{|A|^{1/M}} (\theta - \theta_0)^T \Sigma^{-1} (\theta - \theta_0) + G(\theta) \right] \right\},$$

where A is an $M \times M$ matrix, $|A|^{1/M} \Sigma = A$ and $|\Sigma| = 1$. Note that $|A|^{1/M}$ should be thought of as the covariance scale factor. Then we see that a natural choice for λ is

$$\lambda = \frac{\delta^2}{|A|^{1/M}}, \quad (2.31)$$

i.e. the ratio of the mean squared-error in measurement (market noise) δ^2 to the mean squared-error in our prior knowledge (confidence) $|A|^{1/M}$. A similar conclusion is reached by Golub et al. [75] in their aforementioned paper on linear regression. Fitzpatrick says we should think of the prior knowledge as ellipsoids of constant probability of the prior, with shape and orientation governed by Σ and size given by $|A|^{1/M}$. Similar to our remark for the L-curve analysis, the choice (2.31) for λ essentially ensures that a sensible balance is struck between the measurement error, i.e. confidence, of the observed data and prior knowledge.

2.5 Numerical Examples

Using raw data cited by other papers, we attempt to calibrate the local volatility model. We first calibrate to prices generated using the local volatility surface used in [89]. Hence, it is a synthetic market where the true model is known and of the same class as the model we try to calibrate. We use a Markov Chain Monte-Carlo (MCMC) Metropolis algorithm to sample the posterior distribution of calibrated parameters. Lastly we repeat the procedure using S&P 500 data taken from [28].

2.5.1 Two Datasets

1. We price 66 European call options on the local volatility surface given in [89] (4.4) with 11 strikes and 6 maturities. Similar to [89], we take $S_0 = 5000$, interest rate $r = 0.05$, dividend rate $d = 0.03$. To each of the prices we add Gaussian noise with mean zero and standard deviation 0.1% of the original price [79] and treat these as the observed prices. Coleman et al. [28] do something very similar but take instead Gaussian noise with mean zero and standard deviation 0.02. We take the calibration error acceptance level to be $\delta = 3$ basis points, following the results of [89]. The prices are given in Appendix A.1.

2. We take real S&P 500 implied volatility data used in [28] to price corresponding European call prices and then find a distribution of local volatility surfaces. 70 European call prices are calculated from implied volatilities (using the standard Black Scholes formula (2.12-2.13)) with 10 strikes and 7 maturities. Spot price of the underlying at time 0 is $S_0 = \$590$, interest rate is $r = 0.060$ and dividend rate is $d = 0.0262$. The prices are given in Appendix A.1.

2.5.2 Non-Parametric Discretisation

We represent the local volatility surface $\sigma(S, t)$ by a grid of nodes whose positions are given by $S_{min} = s_1 < \dots < s_j < \dots < s_J = S_{max}$ in the spatial direction and $0 = t_1 < \dots < t_l < \dots < t_L = t_{max}$ in the temporal direction. For each time t_l we construct the unique natural cubic spline through the nodes $(s_1, t_l), \dots, (s_J, t_l)$ to give all values $\sigma(S, t_l)$. Then for $(S, t) \in [s_j, s_{j+1}] \times [t_l, t_{l+1}]$ the value of $\sigma(S, t)$ is found by linear interpolation of the two values $\sigma(S, t_l)$ and $\sigma(S, t_{l+1})$. This is the same approach as in [89].

Assumption 2.5.1. There exist $K, k > 0$ such that, for any surface σ , for at least one of the observed option prices, corresponding to pricing function f say,

$$K > \frac{10^4}{S_0 \delta} \frac{|f(\sigma) - f(\sigma^*)|}{\|\sigma - \sigma^*\|_2} > k.$$

Proposition 2.5.2. *For the local volatility discretisation scheme described above and European call option pricing functions $f_t^{(i)}(\sigma)$ satisfying Assumption 2.5.1, the Bayes estimator $\hat{\sigma}_n(V)$ is consistent.*

Proof. First observe that we satisfy Assumption 2.2.4 since the prior $p(\sigma)$ is continuous at σ^* , $p(\sigma^*) > 0$, the support Σ is closed and bounded (since numerically we truncate the prior Gaussian distribution for small values of $p(\sigma)$) and $p(\sigma)$ is bounded on Σ . So if Assumption 2.5.1 holds for each observation time t then Assumption 2.2.11 is satisfied. Hence, Theorem 2.2.12 applies. \square

For the first example we take nodes positioned on the grid given by

$$\begin{aligned} s &= 2500, 4500, 4750, 5000, 5250, 5500, 7000, 10000, \\ t &= 0.0, 0.5, 1.0, \end{aligned}$$

so $J = 9$, $L = 3$ and there are a total of $M = J \times L = 27$ free parameters (cf 66 calibration prices).

For the second example we take nodes positioned on the grid given by

$$\begin{aligned} s &= 300, 500, 560, 590, 620, 670, 800, 1200, \\ t &= 0.0, 0.5, 1.0, 2.0, \end{aligned}$$

so $J = 8$, $L = 4$ and there are a total of $M = J \times L = 32$ free parameters (cf 70 calibration prices).

After some ordering convention the discrete representation of $\sigma(S, t)$ is given by the parameter vector

$$\theta = \log(\sigma_1, \dots, \sigma_m, \dots, \sigma_M),$$

where $\sigma_{j+(l-1)J} = \sigma(s_j, t_l)$. With this discretisation, the norm function in (2.23) is discretised by

$$\|\log(\sigma) - \log(\sigma_{atm})\|_\kappa^2 \approx (\theta - \theta_{atm})^T C(\theta - \theta_{atm}),$$

where C is the inverse covariance matrix induced by the norm. From the norm given by (2.24), C is non-singular, so for the sake of convention write $A^{-1} \equiv C$. Full detail of how the matrix C is derived is given in Appendix B.1.

Remark 2.5.3. In light of the above, Assumption 2.5.1 is reasonable to make since, in both the simulated and real case, the number of options (66 and 70 respectively) is greater than the number of calibration parameters (the number of nodes are 27 and 32 respectively). Hence, given the longest-dated maturity is equal to the longest-dated node, any surface σ will give prices which differ from those given by σ^* for at least one contract. It is reasonable to expect this difference to be proportional to $\|\sigma - \sigma^*\|_2$.

2.5.3 Metropolis Sampling

Given the price data V , we are interested in the posterior distribution $p(\theta|V)$ given by (2.28). However, because of the high dimensionality of θ , it is unfeasible to analytically find $p(\theta|V)$. Instead, the best we can do is try to draw samples from this distribution and make conclusions based on these samples.

Firstly, observe that the form of the posterior given by (2.28) makes direct sampling slightly difficult. It will be computationally more efficient to first sample from the non-truncated density,

$$g(\theta|V) \propto \exp\left\{-\frac{1}{2\delta^2} [\lambda\|\theta\|^2 + G(\theta)]\right\}, \quad (2.32)$$

and then use the following proposition.

Proposition 2.5.4. Suppose $\{\theta_1, \dots, \theta_n\}$ is a set of samples from $g(\theta|V)$ given by (2.32). Then

$$\{\theta_i : G(\theta_i) \leq \delta^2\}$$

is a set of samples from $p(\theta|V)$ given by (2.28).

Proof. This can easily be verified using the rejection sampling algorithm (see [74] for example) which states the following:

Suppose $\frac{p(\theta|V)}{g(\theta|V)} \leq K$ for all θ and some known constant K . Sample θ at random from density $g(\theta|V)$ and accept θ with probability $\frac{p(\theta|V)}{Kg(\theta|V)}$ as a draw from $p(\theta|V)$. Then an accepted θ has the correct distribution $p(\theta|V)$.

In our case

$$\frac{p(\theta|V)}{g(\theta|V)} = \begin{cases} \frac{1}{\mathbb{P}[\theta \in \Theta : G(\theta) \leq \delta^2]} & \text{if } G(\theta) \leq \delta^2 \\ 0 & \text{if } G(\theta) > \delta^2 \end{cases}$$

so, using the rejection sampling algorithm, each sample θ of $g(\theta|V)$ is accepted with probability 1 if $G(\theta) \leq \delta^2$ and rejected otherwise and each accepted sample has correct distribution $p(\theta|V)$. \square

We now concentrate on generating samples from $g(\theta|V)$. To do this we will use the Markov Chain Monte-Carlo (MCMC) Metropolis algorithm, which proceeds as follows (see [70] for further detail):

1. Select a starting point θ_0 for which $g(\theta_0|V) > 0$.
2. For $r = 1, \dots, n$, sample a proposal $\theta^\#$ from a *symmetric jumping distribution* $J(\theta^\#|\theta_{r-1})$ and

set

$$\theta_r = \begin{cases} \theta^\# & \text{with probability } \min\left\{\frac{g(\theta^\#|V)}{g(\theta_{r-1}|V)}, 1\right\} \\ \theta_{r-1} & \text{otherwise.} \end{cases}$$

Then the sequence of iterations $\theta_1, \dots, \theta_n$ converges to the target distribution $g(\theta|V)$.

To optimise the routine we run m parallel chains, each starting from a different point $\theta_0^{(j)}$ such that the set of starting points $\{\theta_0^{(1)}, \dots, \theta_0^{(m)}\}$ is an *overdispersed* sample of the target distribution. By overdispersed we mean that the samples are more widely distributed than the target distribution (see [71]). We also discard the first b iterations of the run (known as the *burn-in*), since it takes some exploratory time for the algorithm to settle on the target distribution. And we only keep every k th draw from the remaining iterations (known as *thinning*) of the sequence, to reduce the correlation between samples.

Two operations remain to be discussed: the choice of jump function J and how to evaluate the posterior $p(\theta|V)$ for a given θ . Authors such as Beskos & Stuart [12] and Hairer et al. [1] have recommended associating the jump function with a random walk for which the transition kernel is associated with the prior density (2.22). Let A^{-1} be the inverse non-singular covariance matrix introduced in the previous section. By Cholesky decomposition, we can find a matrix B such that $A = BB^T$. Then the jump function $J(\theta'|\theta)$ is given by

$$\theta' = \theta + \sqrt{2du}B\xi,$$

where $\xi \sim N(0, I_M)$ and du is the step size of our random walk. The value of du is chosen so that the acceptance rate of jumps is close to the optimum value of 23% found by Gelman et al. [70].

Finally, to calculate the likelihood value (2.27) for each θ we must price all the European call calibration options, $f_t^{(i)}(\theta)$ for $i \in I_t$, using the model parameter θ . For the local volatility model, we follow the method of [79] of solving the Dupire PDE [51] with appropriate boundary conditions:

$$\begin{aligned} \frac{\partial f}{\partial T} + K(r-d)\frac{\partial f}{\partial K} - \frac{K^2\sigma^2(K,T)}{2}\frac{\partial^2 f}{\partial K^2} &= 0 \\ \forall K \geq 0, \quad f(S, K, 0) &= (S - K)^+, \end{aligned}$$

where $\sigma(K, T)$ is the local volatility surface corresponding to the parameterisation θ . To solve the above PDE we use an implicit finite-difference scheme to give all the prices simultaneously. The weights are chosen beforehand depending on relative volumes likely to be traded (so at-the-money options are weighted more heavily).

The values for n, m, b, k, du for each numerical example are given the Appendix A.2.

2.5.4 Monitoring Convergence of Metropolis Sampling

Convergence can be assured by checking that the *potential scale reduction factor* (PSRF) [70] of estimands of interest, the calibration prices for example, is close to 1 — and at least less than 1.1 in particular. Full details of the calculation and explanation of the procedure are set out below.

Suppose we have found m chains of length n (after discarding the burn-in and using thinning). For each calibration price v we note by v_{ij} the i^{th} draw from the j^{th} simulated chain, for $i = 1, \dots, n$ and $j = 1, \dots, m$. Then we compute the between-sequence variances, B , and within-sequence variances, W , as follows:

$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{v}_{..j} - \bar{v}_{..})^2 \quad \text{and} \quad W = \frac{1}{m} \sum_{j=1}^m s_j^2,$$

where

$$\bar{v}_{..j} = \frac{1}{n} \sum_{i=1}^n v_{ij} \quad \bar{v}_{..} = \frac{1}{m} \sum_{j=1}^m \bar{v}_{..j} \quad s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (v_{ij} - \bar{v}_{..j})^2.$$

Convergence is then assessed by estimating the factor by which the scale of the current distribution for v might be reduced if the length of the chain was allowed to continue in the limit $n \rightarrow \infty$. The potential scale reduction factor (PSRF) is estimated by

$$PSRF(v) = \sqrt{1 - \frac{1}{n} \left(1 - \frac{B}{W} \right)} \tag{2.33}$$

which tends to 1 as $n \rightarrow \infty$. If the PSRF is high, a lot greater than 1.1 for example, then it is likely that continuing simulation will improve inferences based on the target distribution.

The estimate works because B usually overestimates posterior variance assuming the starting distribution is overdispersed. But W usually underestimates the posterior variance because within chain samples have not had sufficient time to range over all the target distribution. But the longer we run the chain, the closer B gets to W and so the closer the ratio B/W gets to 1 and the PSRF estimate for v given by (2.33) gets to 1 also.

See [71] or [70] for further references on PSRF values.

2.5.5 Results

To check the convergence of the Metropolis Sampling we compute the PSRF numbers for the European call prices to which we are trying to calibrate. For the first case, in which we try to calibrate

to 66 noisy European call prices calculated using a known local volatility surface, we give the PSRF values in Table 2.1 below. Recall that values close to 1, and especially ones less than 1.1, imply good convergence and no need for further sampling. By good convergence we mean that the variability of samples within chains is close to the variability of samples between different chains. From Table 2.1 we see that almost all prices have PSRF value 1.0, which indicates excellent sampling. Only some short-dated far out-of-the-money options have values slightly greater than 1.1; but these options had smaller weights w_i in our algorithm so we would expect slower convergence.

Maturity	Strike (units of S_0)										
	0.80	0.90	0.94	0.96	0.98	1.00	1.02	1.04	1.06	1.10	1.20
0.083	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.001	1.003	1.039	1.140
0.167	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.001	1.022	1.156	
0.250	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.012	1.160	
0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.001	1.071	
0.750	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.009	
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.001	

Table 2.1: For the simulated dataset: PSRF values for the calibration call prices (using [89]).

Using the MCMC Metropolis algorithm, 479 surfaces were accepted as samples for the local volatility surface, i.e. gave $G(\theta) \leq \delta^2$ (cf. the 50 found by Hamida & Cont [79]). These surfaces are plotted in Figure 2.1 for the case where prior confidence parameter $\lambda = 1$. Figure 2.1 clearly

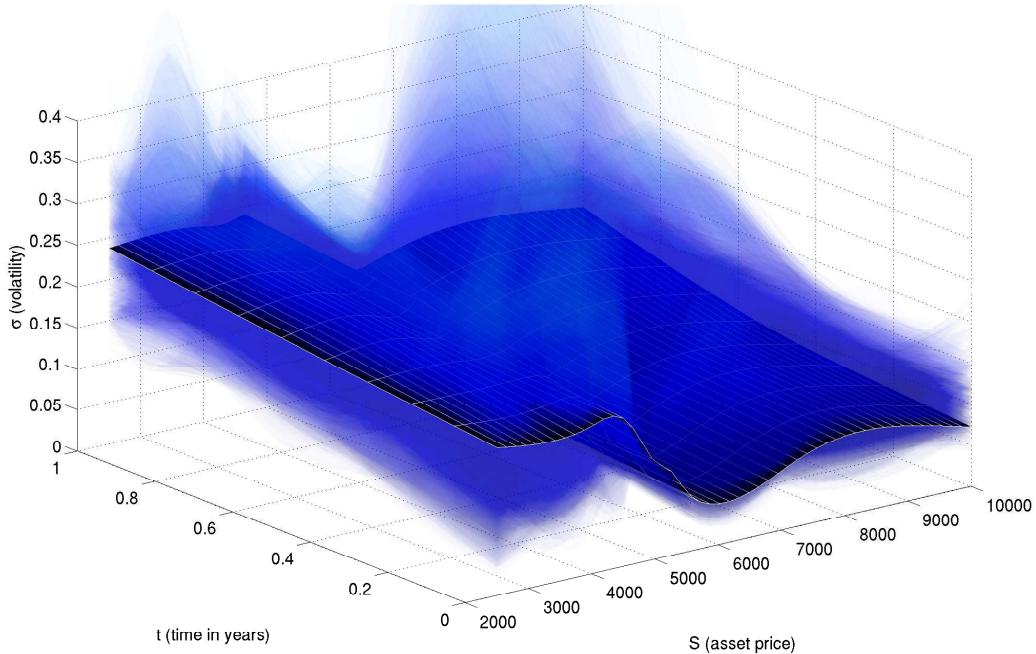


Figure 2.1: For the simulated dataset: using Metropolis Sampling, 479 surfaces from the posterior distribution were sampled and are plotted (where $\lambda = 1$). The true surface is plotted in opaque black.

demonstrates the variety of surfaces which can be calibrated to the same set of prices. We see that, especially for stock values far from S_0 , the volatility becomes very uncertain and varied.

Using this distribution of surfaces, we can construct confidence intervals (or ‘credible sets’ in the language of Bayes) of the value of the local volatility surface $\sigma(S, t)$ at any point (S, t) . Figure 2.2 shows the 95% and 68% pointwise confidence intervals. Figure 2.2 shows that, close to the spot

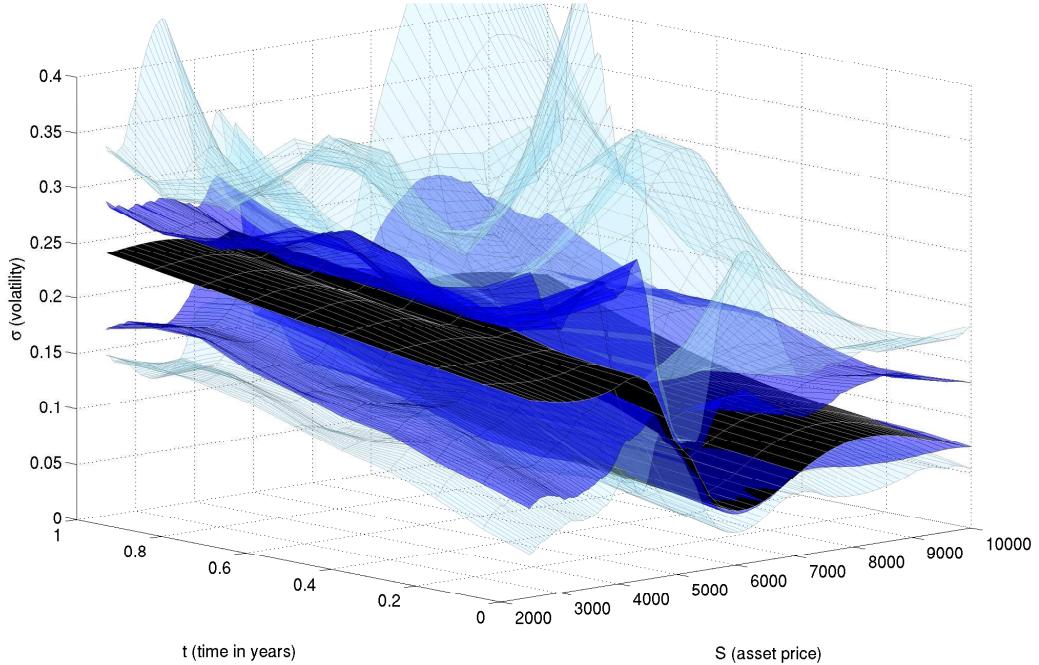


Figure 2.2: For the simulated dataset: using the 479 sampled surfaces (where $\lambda = 1$), the confidence intervals are found pointwise. The true surface is plotted in opaque black.

price S_0 , the bounds are very tight, but the further away-from-the-money we go, the looser these bounds become. However, we see that the true surface is almost captured within the 68% confidence interval and completely captured by the 95% confidence interval. This is a strong result and could, for example, be used to find lower and upper volatility bounds for implementation of the uncertain volatility model studied by Lyons [102] and Avellaneda et al. [5], where the instantaneous volatility σ_t is assumed to lie within an interval $[\sigma_t^l, \sigma_t^u]$.

This distribution of local volatility surfaces can be used to price other contracts. In particular, we are interested in the probability density of prices implied by the Bayesian posterior distribution of local volatility surfaces, the true value as priced on the correct surface, the bid-ask spread (calculated as the true price $\pm \delta$ basis points), the value calculated on the surface which gave the smallest regularised calibration error — i.e. the maximum a posteriori surface (MAP), and the Bayesian

model average price, given by

$$\frac{1}{N} \sum_{i=1}^N f(\theta_i) \approx \int_{\theta} f(\theta) p(\theta|V) d\theta, \quad (2.34)$$

where f is the pricing function and $\theta_1, \dots, \theta_N$ are the surfaces found by Metropolis sampling. Note that, because the parameters $\theta_1, \dots, \theta_N$ are *samples*, the Bayesian weighting of each in the sum (2.34) is $1/N$ rather than $p(\theta_i|V)$. The other quantities are obtained as follows. The Bayesian posterior probability density of prices is simply the histogram corresponding to $f(\theta_1), \dots, f(\theta_N)$ (since they have equal weighting in the sum (2.34) above). The true price $f(\theta^*)$ is found by pricing on the true surface θ^* and the bid-ask spread estimated by $[f(\theta^*) - \delta S_0/10^4, f(\theta^*) + \delta S_0/10^4]$. The MAP price is taken to be $f(\theta^{MAP})$, where θ^{MAP} is the sample that has greatest posterior density i.e. $p(\theta^{MAP}|V) \geq p(\theta_i|V)$ for all $i = 1, \dots, N$. Note that the MAP price does not correspond to the maximum of $f(\theta_i)$. We noted in Remark 2.4.1 that the MAP parameter corresponds to the solution of classical Tikhonov regularised error function minimisation problems, hence we plot the MAP price for aid of comparison to the Bayes price.

Remark 2.5.5. It is necessary to point out that the MAP estimator we use will not be the *true* MAP estimator θ^{MAP} in the sense that it is not the surface which maximises the posterior density (2.28). We have optimally sampled (2.28), but not optimised it in any other sense. However, we observe that our MAP estimator gives a weighted average basis-point calibration error (2.17) of 1.84 for 66 prices, compared with the 2.65 for 10 prices achieved by Jackson et al. [89]. Hence, our density sampling has found a MAP surface which gives optimisation comparable to that found by papers using different optimisation routines.

In the following two figures we price an up-and-out barrier call option and an American put option. These claims are path-dependent and, as such, much more sensitive to changes in the local volatility surface. In both graphs we plot the Bayesian posterior probability density of prices, true, MAP, Bayes prices and estimated bid-ask spread of the true price. Figure 2.3 and Figure 2.4 show that the surface with the smallest regularised calibration error can lie many basis points away from the true price. The Bayesian price on the other hand reflects the entire distribution and the incorporated prior information (i.e. regularisation) to give a much closer price, which lies well within the bid-ask spread.

We can also look at verifying the consistency property stated in Proposition 2.5.2. We simulated a path for 12 weeks on the true surface and priced the same 66 European call options at the start of each week. We recalculated the error functional $G(\theta)$ and updated the Bayesian posterior. Note that this is very different from the single calibration procedure described before. We are now iterating the calibration and updating the Bayesian posterior to incorporate the new observed prices every week, so there are 12 calibrations rather than 1. The results of the change in Bayesian posterior

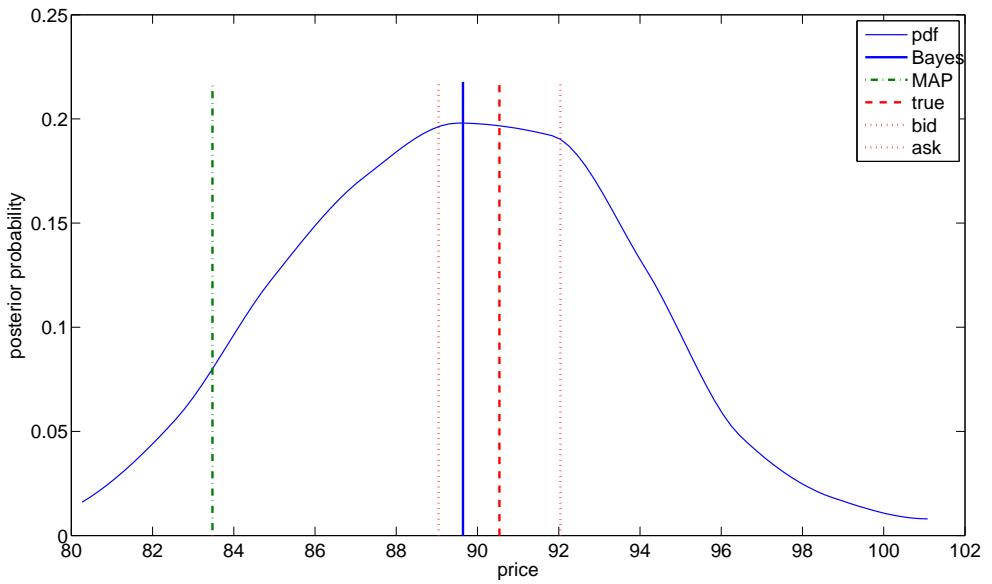


Figure 2.3: For simulated dataset: prices for up-and-out barrier call option strike 5000 ($S_0 = 5000$), barrier 5500 and maturity 3 months. Included are the true price (found on the true surface) with its bid-ask spread, the MAP price, and the Bayes price with its associated posterior pdf of prices.

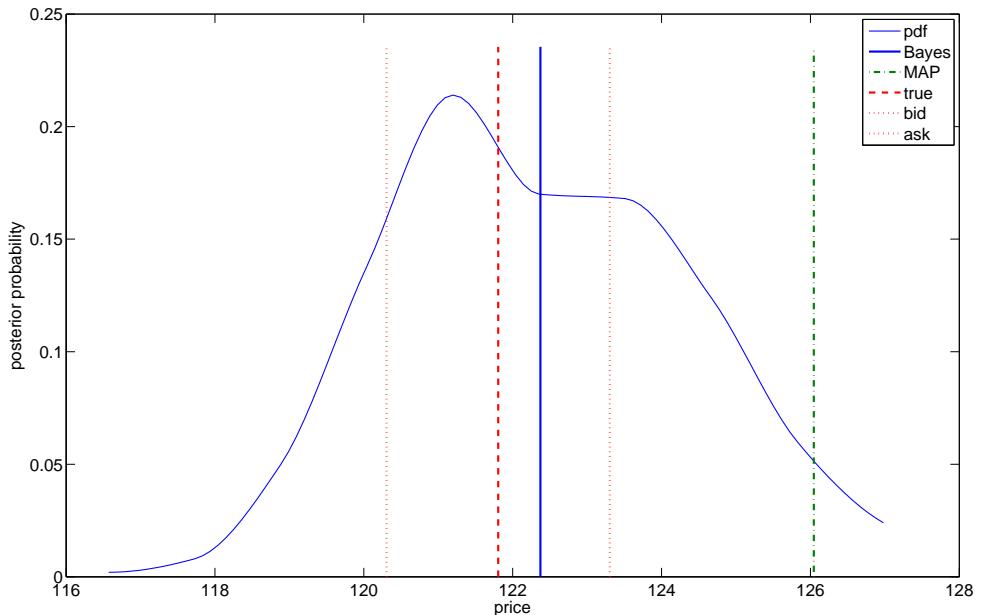


Figure 2.4: For the simulated dataset: prices for American put option with strike 5000 ($S_0 = 5000$) and maturity 1 year. Included are the true price (found on the true surface) with its bid-ask spread, the MAP price, and the Bayes price with its associated posterior pdf of prices.

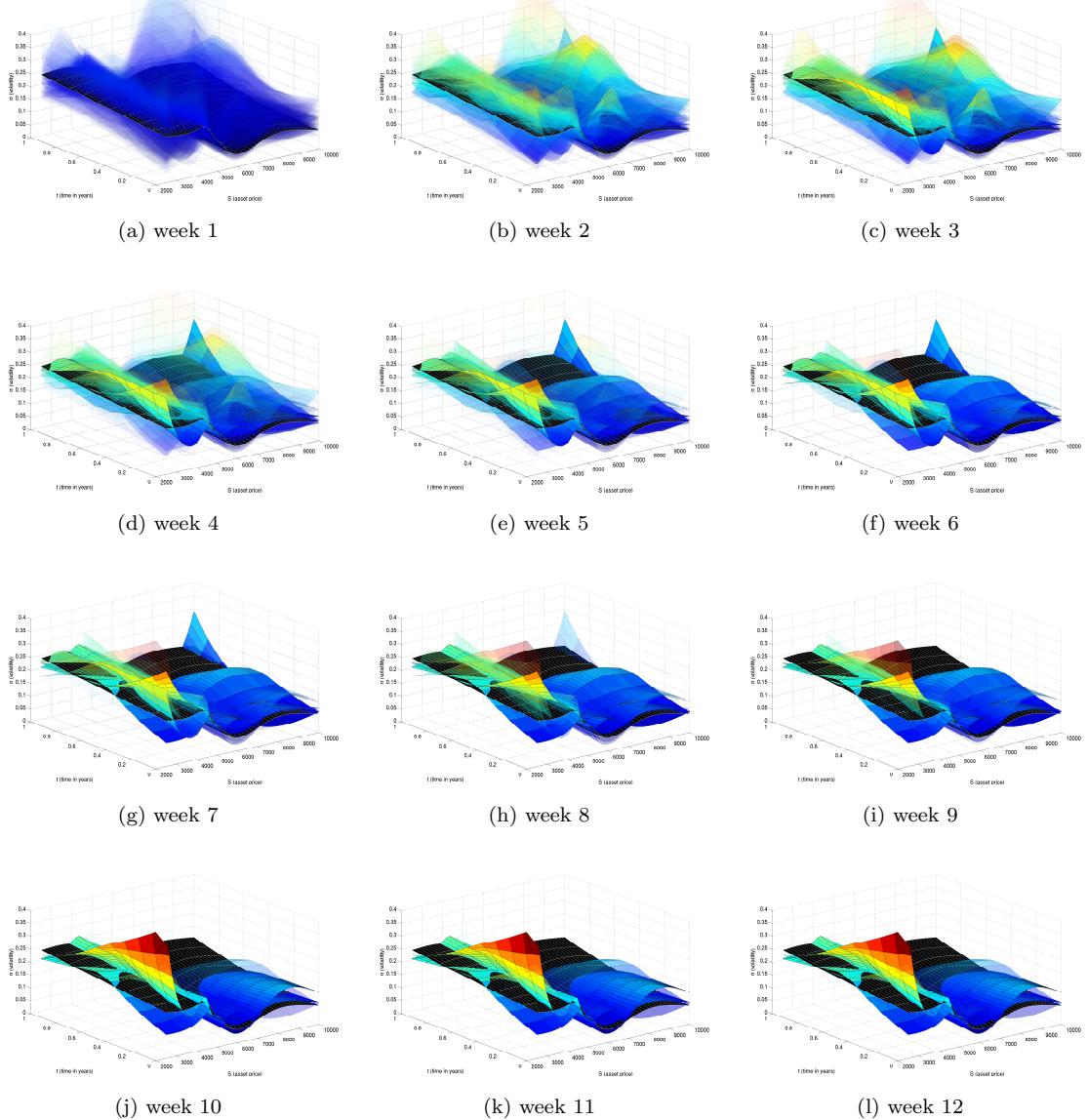


Figure 2.5: For the simulated dataset: a path is simulated on the true local volatility surface and the Bayesian posterior is updated using the newly observed prices each week for 12 weeks. The transparency of each surface reflects the Bayesian weight (see main text) of the surface. The true surface is plotted in opaque black.

are given in Figure 2.5. Full details of how the Bayesian posterior is updated is given later in the thesis in Section 4.4.4 of Chapter 4, where we test optimal Bayesian hedging strategies. In essence, we implement importance sampling by using the new error function to update the posterior value of the surfaces we have already sampled via the Metropolis algorithm. Note that the transparency of each surface in Figure 2.5 still corresponds, not to the posterior density of the surface, but to the weight in the Bayesian mean (2.34). It is worthwhile clarifying this point. In the Bayesian mean calculation (2.34) at the first calibration time, the ‘Bayesian weights’ are $y_0^{(i)} = 1/N$ for each surface θ_i for $i = 1, \dots, N$ — hence in Figure 2.1 all plotted surfaces have the same degree of transparency. However, after the first recalibration, the new Bayesian mean calculation for a function f will be

$$\sum_{i=1}^N y_1^{(i)} f(\theta_i),$$

for some Bayesian weights $y_1^{(1)}, \dots, y_1^{(N)}$ summing to 1. How these weights are found is detailed in Section 4.4.4. However, the weights are no longer equal. And to reflect this in Figure 2.5 we have varied the transparency of the plotted surfaces to reflect the weight. So, for example, a surface with greater Bayesian weight will be more opaque.

Figure 2.5 shows that after about 5 weeks, the Bayesian posterior has settled and only a handful of surfaces have significant weight. Moreover, these surfaces are close to the true surface (plotted in opaque black). At recalibration time t_k , say, the section of the local volatility surface corresponding to $\{\sigma(S, t) : 0 \leq t < t_k\}$ no longer contributes to the observed calibration prices, so the section $\{\sigma(S, t) : 0 \leq t < t_k\}$ of some of the heavily weighted surfaces is very different to the true surface. This is especially noticeable in the tails for very small S and very large S . We must remember that we have only sampled the Bayesian posterior, and hence if one of our samples is not the true surface (which it is not) then we will never settle on this true surface, but settle on the closest few, as Figure 2.5 does. Instead of updating the Bayesian weights of the surfaces we found at time 0, we could resample the Bayesian posterior each week or each day using the Metropolis algorithm. But observe that, for the first re-sampling, we are calibrating to $66 + 66 = 132$ prices and 198 prices on the second re-sampling, which makes sampling much harder and takes longer to converge. Nevertheless, using the proxy updating procedure we have, we still see a clear sense of convergence to the true surface.

For the second case, in which we try to calibrate to 70 S&P 500 European call prices, we give the PSRF values in Table 2.2 below. Values close to 1, and especially less than 1.1, imply good convergence and no need for further sampling. Again we see that most options have PSRF value 1.0, indicating that the Metropolis sampling routine has been allowed to run for a sufficient time. Only some short-dated far out-of-the-money calls have have values just above 1.1 which is again attributable to their smaller weights w_i in the error function (2.17).

Maturity	Strike (units of S_0)									
	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20	1.30	1.40
0.175	1.000	1.000	1.000	1.000	1.002	1.022	1.105	1.177	1.166	1.119
0.425	1.000	1.000	1.000	1.000	1.000	1.001	1.025	1.130	1.218	1.169
0.695	1.000	1.000	1.000	1.000	1.000	1.000	1.001	1.015	1.123	1.188
0.940	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.001	1.018	1.106
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.011	1.088
1.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.002	1.022
2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.002

Table 2.2: For S&P 500 dataset: PSRF values for the calibration call prices (using [28]).

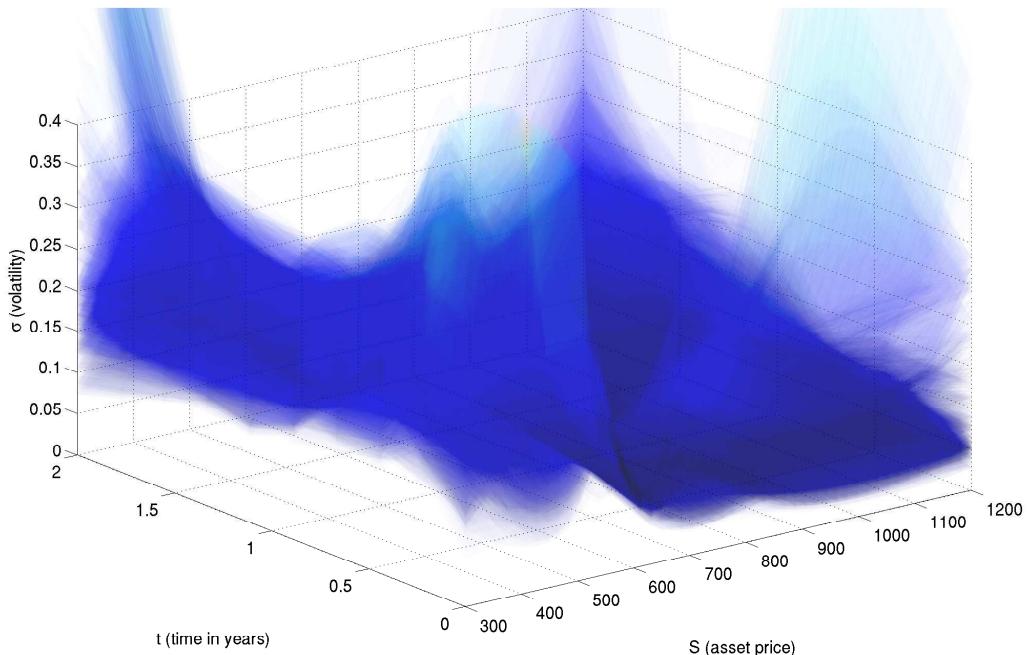


Figure 2.6: For S&P 500 dataset: using Metropolis Sampling, 600 surfaces from the posterior distribution were found and are plotted (where $\lambda = 1$).

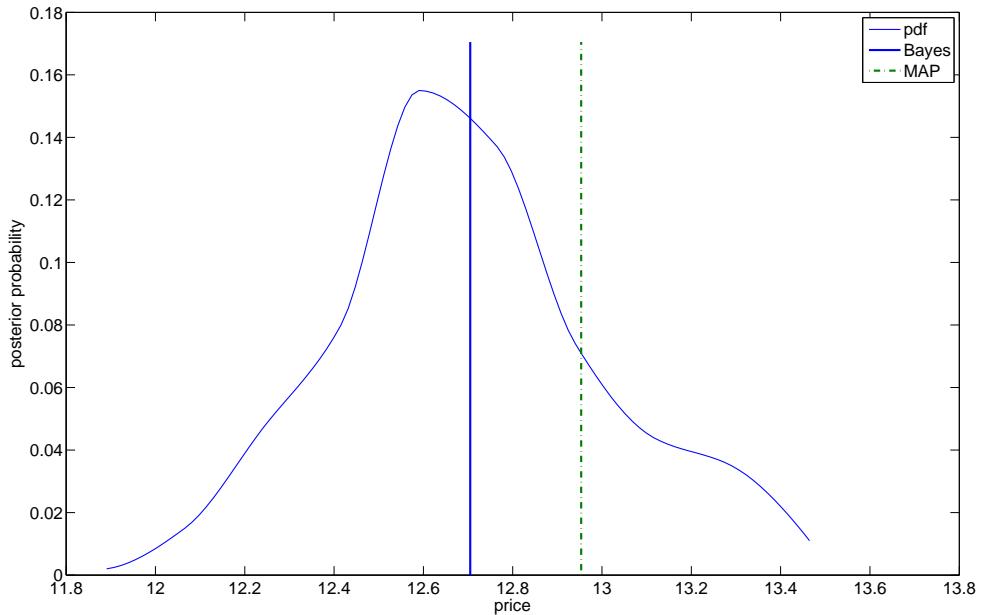


Figure 2.7: For S&P 500 dataset: prices for American put option with strike \\$590 ($S_0 = \590) and maturity 1 year. Included are the MAP price and the Bayes price with its associated posterior pdf of prices.

Figure 2.6 gives a plot of 600 samples from the posterior (so satisfied $G(\theta) \leq \delta^2$, this time for $\delta = 4.5$ basis points). Again we see that, especially in the wings and for short times, the spread of volatilities is enormous.

The second graph, Figure 2.7, shows the distribution of prices for an American put option written on the same underlying corresponding to the Bayesian posterior distribution. Again the MAP and Bayes prices are marked on the graph. However, in this case we have used real data so we do not know the real local volatility surface (or even if one exists) so cannot compare with true results. Nevertheless, we see that the Bayes price is closer (by design) to the centre of the distribution of prices whereas the MAP price is closer to the tail of the distribution.

2.5.6 Robustness

Recall that the Bayesian posterior we calculate largely depends upon the form of the prior and its regularisation properties. Also recall Hadamard's third criterion for well-posedness from Section 2.1.2: iii) the solution depends continuously on the data. In this section we conduct some robustness (also referred to as *sensitivity analysis* in the literature, e.g. Berger [11]) tests on the datasets of the examples studied in the previous section to quantify how much the solution varies with respect to both changes in the form of the prior and the observed prices.

In his book on robustness, Huber [84] defines *robustness* as signifying

$$\text{insensitivity to small deviations from the assumptions,} \quad (2.35)$$

where the assumptions might refer to the form of the prior or the accuracy of the observations or some other factors. Alternatively, Aster et al. [3] refer to *robust estimation procedures* as techniques that are resistant to *outliers*, which are data points wildly discordant with the mathematical model. In our case the mathematical model is a financial one and outliers are unlikely in this context because a liquid option price wildly discordant with a calibrated model would introduce an obvious arbitrage opportunity which is rare and, if it did occur, would quickly evaporate as heavy trading brought its price to normality. Hence, for the purposes of this thesis we will use (2.35) as our criterion for robustness.

The literature on robust methods for calibration of the local volatility model is fairly sparse. However, some theoretical work has been carried out by Samperi [119] on the calibration via relative entropy minimisation method introduced by Avellaneda et al. [4]. Samperi regularises the method by introducing bounds on the local volatility function (in the manner of uncertain volatility) and then solves a stochastic optimal control problem for value function $W(s, t; \lambda)$ corresponding to the infimum of a regularised error function of the form (2.18). Samperi shows that W is twice continuously differentiable with respect to λ and strictly convex and hence depends continuously on the observed prices.

In a similar vein, Cont & Tankov [33] show robustness of their method for recovering a general Lévy process from observed calibration prices. They also use relative entropy regularisation and show insensitivity of the solution to small changes in the data and choice of prior. [33] shows that, as the noise δ of the observed prices vanishes, the solution of the regularised minimum entropy problem converges to that of the original minimum entropy problem. Similarly, they show that, for any solution of the regularised minimum entropy problem with prior P , as $P \rightarrow P'$, this solution has a weakly convergent subsequence converging to the solution of the calibration problem with prior P' .

To numerically assess criterion (2.35) for our Bayesian calibration method two questions need to be answered: how do we create ‘small deviations from the assumptions’ and how do we measure the ‘insensitivity’. For the first question we adjust the form of the prior by changing the value of κ defined in (2.24) and change the observed prices by adding Gaussian noise e with different values of standard deviations ε . In each case we hold all other parameters fixed, in particular the confidence parameter at $\lambda = 1$.

For the question of measurement of the effects of the changes, we calculate two things. Firstly, we recalculate the distribution of prices for the same two options: American put option and up-and-out barrier call option. Secondly we calculate the ‘distance’ $d(p, q)$ between the new Bayesian posterior $q(\theta|V)$ and the original Bayesian posterior $p(\theta|V)$ found in the previous section. There are

a number of well-known functions d commonly used to measure the distance between two probability measures e.g. the Lévy distance, the Prohorov metric, the Bounded Lipschitz metric (see Huber [84] for details). However, for the purposes of our tests, we will use the distance studied by Matusita, in his papers [104] and [105], for two-sample problems such as ours:

$$d(p, q) = \left[\int_{\Theta} \left(\sqrt{p(\theta|V)} - \sqrt{q(\theta|V)} \right)^2 d\theta \right]^{1/2}. \quad (2.36)$$

Note that we have a straightforward upper bound of $\sqrt{2} \approx 1.4$ for d . To compute (2.36) using our new sample of N surfaces $\theta_1, \dots, \theta_N$ we use the Bayesian updating formula described by equation (4.28) in Chapter 4 Section 4.4.4:

$$\begin{aligned} d(p, q) &\approx \left[\sum_{i=1}^N \left(\sqrt{\frac{1}{N} \frac{p(\theta_i|V)}{q(\theta_i|V)}} - \sqrt{\frac{1}{N}} \right)^2 \right]^{1/2} \\ &= \left[\frac{1}{N} \sum_{i=1}^N \left(\sqrt{\frac{p(\theta_i|V)}{q(\theta_i|V)}} - 1 \right)^2 \right]^{1/2}. \end{aligned} \quad (2.37)$$

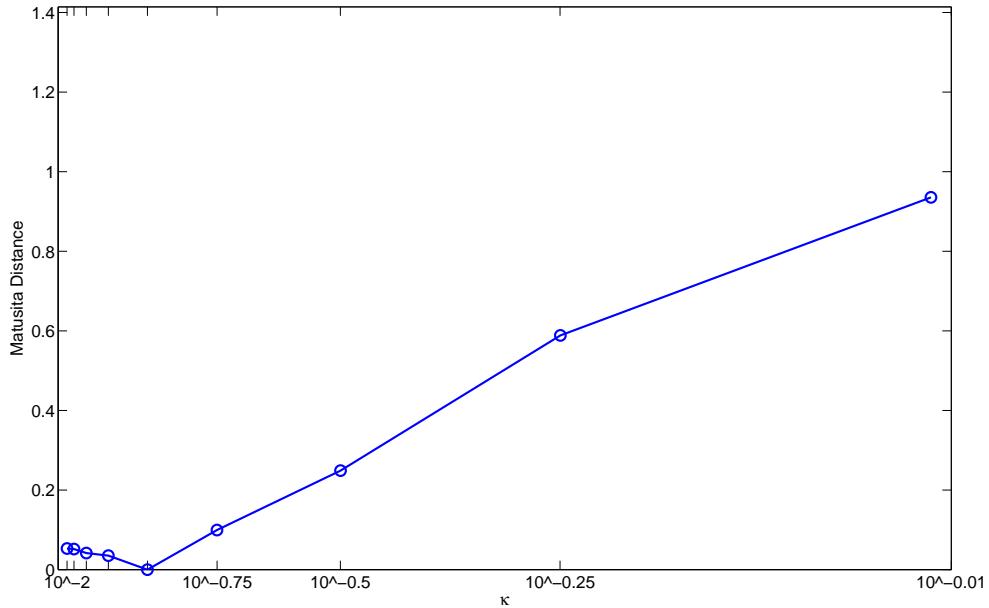


Figure 2.8: Matusita distance between original posterior probability measures and posterior probability measure found for different values of κ . The Matusita distance (2.36) is calculated using formula (2.37)

We first plot the Matusita distances given by (2.37) for different values of κ in Figure 2.8. The value of κ in the previous section's experiments was $\kappa = 10^{-1.0}$ and in the graphs below we repeat the experiments for

$$\kappa \in \{10^{-2.00}, 10^{-1.75}, 10^{-1.50}, 10^{-1.25}, 10^{-1.00}, 10^{-0.75}, 10^{-0.50}, 10^{-0.25}, 10^{-0.1}\}.$$

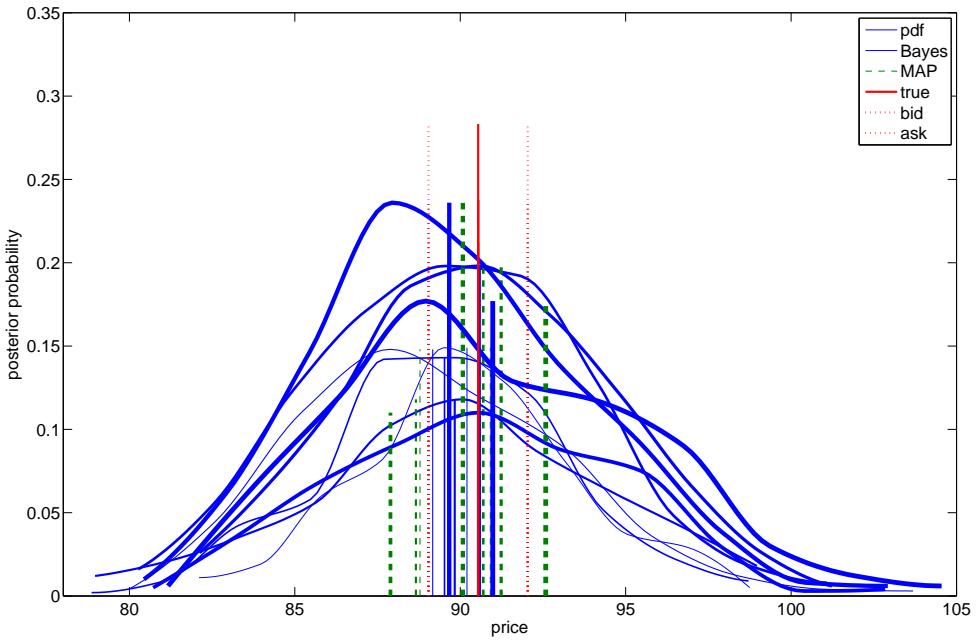


Figure 2.9: Prices for up-and-out barrier call option strike 5000 ($S_0 = 5000$), barrier 5500 and maturity 3 months for different κ . Included is true price with its bid-ask spread, MAP price, and Bayes price with Bayesian posterior pdf of prices. The thickness of the lines increases with κ .

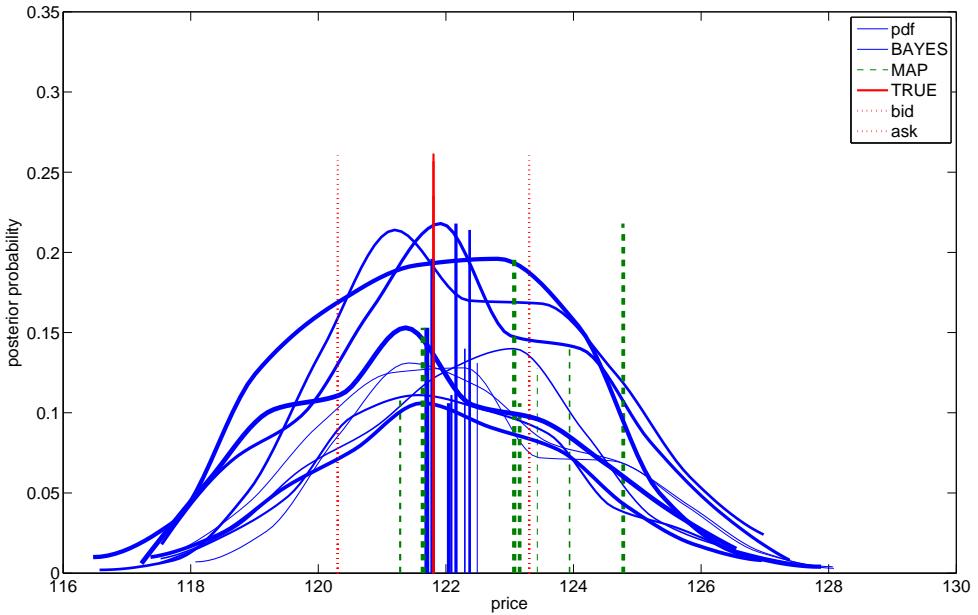


Figure 2.10: Prices for American put option with strike 5000 ($S_0 = 5000$) and maturity 1 year for different values of κ . Included is the true price with its bid-ask spread, the MAP price, and the Bayes price with Bayesian posterior pdf of prices. The thickness of the lines increases with κ .

In Figure 2.9 and Figure 2.10 we plot the graphs equivalent to Figure 2.3 and Figure 2.4 respectively of the previous section, but for different values of κ . Both Figure 2.9 and Figure 2.10 show that the distribution of prices is fairly robust to changes in κ since the peaks are roughly at the same prices and the tails exhibit similar behaviour. Even more encouraging is the robustness of the Bayes estimate which lies in the bid-ask spread for both the barrier and American option for all values of κ . This is in contrast with the MAP estimate, which tends to fall outside of the bid-ask spread for larger values of κ . The Matusita distances in Figure 2.8 exhibit linear behaviour, which is a good sign. Observe that for $\kappa = 10^{-1}$ we recover the base case, so the distributions are the same and the distance between them is zero. Recall that we wanted to check ‘insensitivity to small deviations from the assumptions’. Linearity implies that a small deviation in the prior does not cause a disproportionately large change in the consequent distribution, so we seem to have a strong case for arguing that the Bayesian method is robust with respect to changes in the prior.

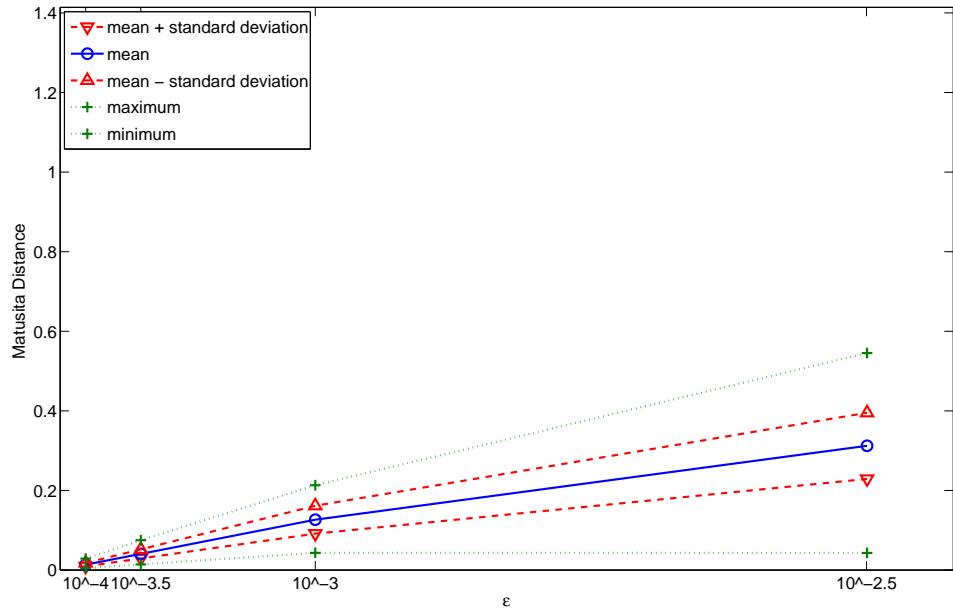


Figure 2.11: Matusita distance between original posterior probability measures and posterior probability measure found for different values of ϵ . There are 100 repetitions for each value of ϵ and hence we have plotted the mean, the mean \pm standard deviation and minimum and maximum distance for each ϵ . The Matusita distance (2.36) is calculated using formula (2.37)

Next we plot the graphs for the same experiments but for the case where we hold $\kappa = 10^{-1}$ fixed and instead add Gaussian noise e with standard deviation ϵ to the observed market prices quoted in Appendix A.1. We consider noise levels of

$$\epsilon \in \{10^{-4.0}, 10^{-3.5}, 10^{-3.0}, 10^{-2.5}\}$$

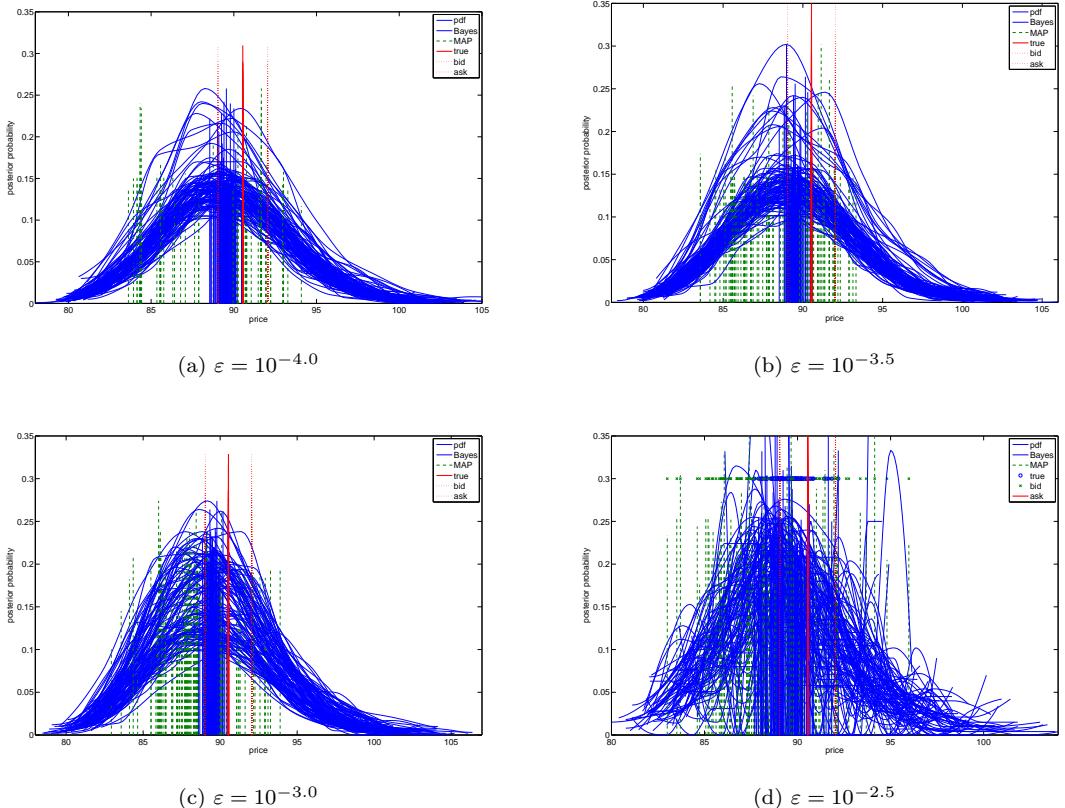


Figure 2.12: Prices for up-and-out barrier call option with strike 5000 ($S_0 = 5000$), barrier 5500 and maturity 3 months. Each graph corresponds to a different value of ε and shows the estimators for 100 different noise additions. Included is the true price (found on the true surface) with its bid-ask spread, the MAP price, and the Bayes price with its associated posterior pdf of prices.

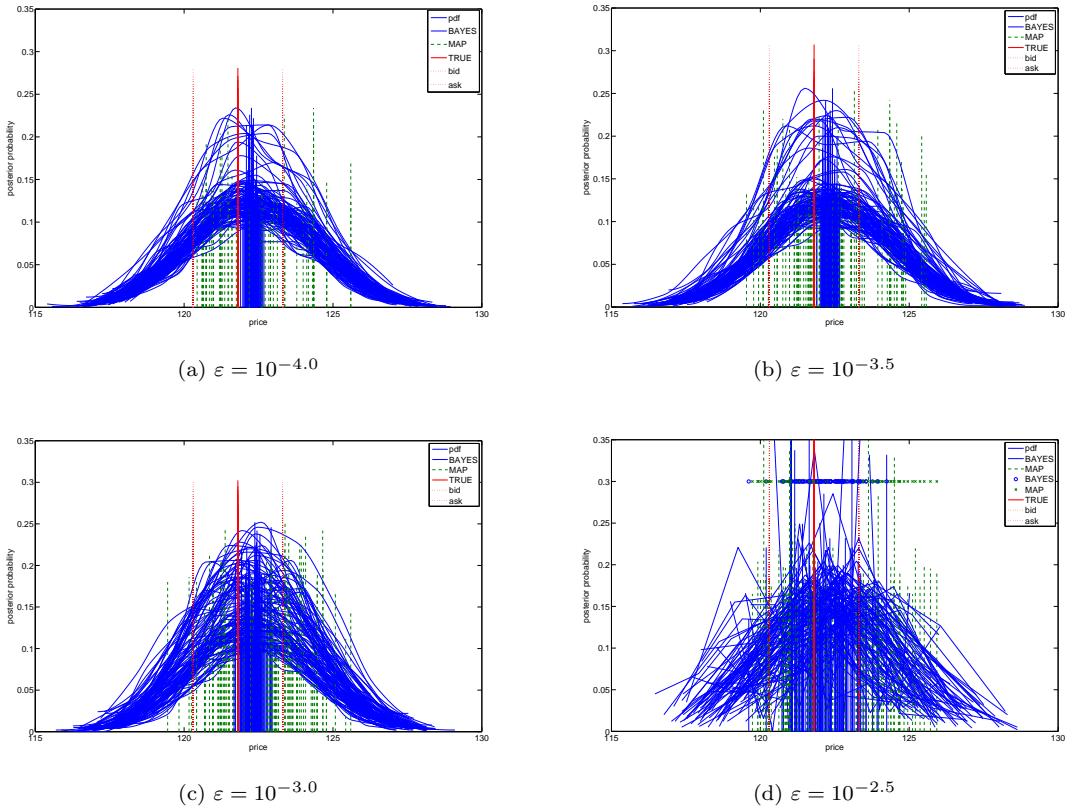


Figure 2.13: Prices for American put option with strike 5000 ($S_0 = 5000$) and maturity 1 year. Each graph corresponds to a different value of ε and shows the estimators for 100 different noise additions. Included is the true price (found on the true surface) with its bid-ask spread, the MAP price, and the Bayes price with its associated posterior pdf of prices.

and run the calibration procedure for 100 different noise additions for each ε . Again we first plot the Matusita distances for the 100 fixed noise realisations corresponding to (2.37) for the different values of ε in Figure 2.11. Then, for each value of ε we plot the 100 MAP and Bayes estimates of the price and posteriors for the same barrier and American options used previously. The graphs are given in Figure 2.12 and Figure 2.13.

The graphs for the barrier option and American option are largely similar and exhibit the same features. First of all we see that as the noise is increased the closeness of the distributions of prices deteriorates and for $\varepsilon = 10^{-2.5}$ few surfaces have been calibrated so the distributions become non-smooth and irregular. The MAP prices are even more sensitive to the noise and can miscalculate the price by up to 10-15%. In stark contrast, the Bayes prices prove to be very robust. For the barrier option for $\varepsilon = 10^{-3.5}$ only one out of a hundred Bayes estimates is slightly outside the bid-ask spread and for $\varepsilon = 10^{-2.5}$ only a handful of Bayes estimates fall slightly below the bid price. The distances in Figure 2.11 tell a similar story to before. The average change is approximately linear with the noise added to the observed market prices. Again we conclude that, unlike the MAP estimate, the Bayes estimates are robust to changes in the observed prices.

2.6 Extensions

In this chapter we have introduced the Bayesian framework for calibrating finite-dimensional parameters of financial models to market prices. Analytic results have been found that prove the Bayesian estimate for the true model parameters is consistent over time when an underlying constant parameter exists. We then demonstrated a practical method for formulating the prior and likelihood functions necessary for the Bayes procedure and applied it to the case of the local volatility model. Numerical examples were presented and showed the improvement in pricing of the Bayesian procedure over common MAP estimates.

Although we have concentrated on the local volatility model throughout the second half of the chapter, the methodology provided in Section 2.4 and Section 2.5 is very general and can be applied to any parametric or non-parametric model. The empirical investigation by Carr et al. [22] led them to conclude that processes for indices and stocks tend to be pure jump processes of infinite activity and with finite variation. Hence, we further demonstrate our method through an application to the more general non-parametric jump diffusion model below.

We note in passing that there is a wealth of literature on calibrating jump-diffusion processes and more general Lévy processes. Relative entropy minimisation methods are used by Cont & Tankov [31], [33]. More recently, Belomestny & Reiss [9] have used spectral calibration of non-parametric exponential Lévy models. They directly invert the option pricing formula using Fourier transforms and use truncation of higher frequencies as their regularisation method. Nothing is mentioned, however, of robustness of these inversions to deviations in the model assumptions.

Example 2.6.1. (*Jump Diffusion*). We follow the model setup of Cont & Tankov [31]. Let X be a compound Poisson process (so finite activity) with characteristic triplet (σ^2, γ, ν) . Then it has Lévy-Khinchin representation

$$\phi(z) = \mathbb{E}[e^{izX_t}] = \exp \left\{ t[-\frac{1}{2}\sigma^2 z^2 + i\gamma z + \int_{-\infty}^{\infty} (e^{ix} - 1)\nu(x)dx] \right\}. \quad (2.38)$$

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{Q})$ be a filtered probability space. Take S to be a \mathbb{Q} -martingale which follows a jump-diffusion process:

$$S_t = e^{X_t}.$$

Since S is a martingale, $\phi(-i) = 1$, so the drift

$$\gamma = -\frac{1}{2}\sigma^2 - \int_{-\infty}^{\infty} (e^x - 1)\nu(x)dx$$

and the only remaining unknown parameters are σ^2 and ν .

We tackle this calibration problem in an identical fashion to the local volatility model. First of all we model σ as a scalar constant and ν to be of finite activity (so $\int \nu(x)dx < \infty$). We then observe the characteristics we would expect of the two parameters that will be recast into the Bayesian prior:

Positivity: the volatility σ should be positive and the density function $\nu(x)$ should of course be positive.

Monotonicity: the density function $\nu(x)$ should be maximum at $x = 0$ and monotone either side; we should expect that either side of zero, the larger the jump size is (be it negative or positive), the less frequently they occur.

Consistency: for small values of t especially, σ should be close to today's at-the-money (ATM) and ν should be close to yesterday's density function ν_0 .

In view of these assumptions we take as our prior for the jump diffusion model the function

$$p((\sigma, \nu)) \propto \exp \left\{ -\frac{1}{2\delta^2} \lambda \left[\|\log(\sigma) - \log(\sigma_{atm})\|_{\iota}^2 + \|\log(\nu) - \log(\nu_0)\|_{\vartheta}^2 \right] \right\},$$

where the norms on the right hand side are given by

$$\begin{aligned} \|u\|_{\iota}^2 &= \iota |u|^2, \\ \|u\|_{\vartheta}^2 &= (1 - \vartheta) \|u\|_2^2 + \vartheta \int_{-\infty}^0 |\frac{d^2 u}{dx^2}|^2 dx + \int_0^{+\infty} |\frac{d^2 u}{dx^2}|^2 dx. \end{aligned}$$

Again, ι and ϑ are pre-specified constants and λ depends on how confident we are in our prior. The positivity assumption is satisfied by working in the log space and the consistency assumption is achieved through $\iota |u|^2$ and $\vartheta \|u(x)\|_2^2$. The monotonicity assumption is accounted for by minimising the integral of the second derivative of the log-density either side of zero.

In Section 2.5.6 we carried out some sensitivity analysis of the results of Section 2.5 to changes in the value of κ and ε to judge robustness of the prior and data respectively. There are, however, a variety of other ways we could have tested robustness. For example, we could calculate how robust the results are to non-Gaussian priors. To this end, Berger [11] suggests using an ‘ ϵ -contamination class’ of priors

$$\Gamma = \{\pi : \pi(\theta) = (1 - \epsilon)p(\theta) + \epsilon q(\theta), q \in \mathcal{Q}\}$$

for $\epsilon \in (0, 1)$ and set of allowed contaminations \mathcal{Q} to test the assumption of a Gaussian prior distribution. And we could look at the similar concept of an ‘ ϵ -contamination class’ of the data. Alternatively we could measure the sensitivity of the option-price distributions to absence of one or more of the observed calibration prices. This would reveal whether the results are disproportionately dependent on a few data points. Equally we could add extra prices, possibly erroneous (introducing arbitrage), into set of calibration prices and measure the effect on the distribution of prices of other options (see Huber [84] for further details).

Having found the Bayesian posterior density $p(\theta|V)$ there is a variety of further useful analysis that we can do. A natural thing to do would be to use the posterior to derive a measure for the model uncertainty of a contract. For any payoff, a distribution of prices can be found (as we showed in Section 2.5.5 for barrier and American options), and this distribution can be used to assign a model uncertainty value to the contract. Such measures would be important for a risk manager and for an agent trying deciding how to invest in different products. This idea is further developed in Chapter 3.

Another, perhaps more important, use of the Bayesian posterior would be to use it to develop better hedging strategies. This is more fundamental than pricing as typically a trader will be more interested in the hedging strategy (the cost of which will then correspond to the trader’s price for the contract) than a stand-alone price. To this end, the Bayesian loss functions introduced in Section 2.1.4 could be designed to correspond to hedging losses so that the Bayes estimator is the parameter θ which minimises the expected hedging loss. This idea is further developed in Chapter 4.

Chapter 3

Model Uncertainty

The pricing of complex derivatives is sensitive to changes in the pricing model and hence the study of model uncertainty has become more important, especially in industry. In this chapter we consider measures of the uncertainty in option pricing models. Motivated by Cont's [30] analogy with coherent and convex market risk measures, we define axiomatic setups for both coherent and convex model uncertainty measures. In both cases we find representation theorems analogous to the market risk representations of Artzner et al. [2] and Fritelli & Gianin [69]. We then use Bayesian posterior distributions with these representations to construct a wide variety of informative model uncertainty measures.

3.1 Introduction

Over the previous 30 years, the volume and variety of financial derivatives traded have increased dramatically. Correct pricing and hedging of American options, barrier options and complex instruments such as credit derivatives and volatility derivatives has become paramount. Recent events in the credit markets in particular have demonstrated how devastating improper credit assumptions and mis-pricings can be. A little over a decade ago, research on model risk began to gather momentum as mathematicians such as Derman [44], Kerkhof [95] and Hull & Suo [86] began clarifying model risk and observers like Elliott [56] commented on the substantial losses corporations were making because of mis-pricings.

As discussed in Chapter 2, calibrating a model to market prices is often very difficult for a number of reasons including an insufficient number of observable prices and lack of robustness. This becomes particularly troublesome when we try to compute the corresponding hedging strategies. For example, Figlewski and Green [77] backtested historical data assuming a Black-Scholes model and found sizeable pricing and hedging errors arose because of incorrect volatility estimation. And more recently Cont [30] has investigated the large hedging losses that can result from incorrect modelling of the underlying.

Also discussed in the previous chapter was that much of the literature on calibration methods only focuses on finding the single best-fit parameter (be it a scalar or a finite-dimensional vector or a function) for a given set of market prices. Inadequate attention is paid to measuring how robust the estimate is or whether other parameters can reproduce market prices equally/sufficiently well. This is a major shortcoming in the current literature. A measure of the uncertainty of the calibration parameter is vital for two reasons. Firstly, it gives a good indication of how suitable our model and calibration dataset are. Secondly, it enables better risk management and provides clear quantitative measures of potential pricing errors, hedging losses, and other important financial quantities. This is of particular importance in light of recent revisions by the Bank for International Settlements' Basel Committee on Banking Supervision, [112] and [111], which explicitly put increased emphasis on operational risk, which includes model risk.

The aim of the chapter is to construct measures of model uncertainty using a Bayesian framework. Not only does a Bayesian approach allow us to incorporate investors' sentiments and prior beliefs in model parameters but it also enables us to extract an entire distribution of calibrated parameters rather than a single best-fit one. This approach is in a similar spirit to authors such as Lyons [102] and Avellaneda et al. [5] who have studied uncertain volatility models where the unknown volatility is assumed to lie within an interval. However, we aim to extract more information and actually put a distribution on this interval for the unknown parameter.

For our Bayesian approach we generalise the axioms and formulation of Cont [30] and propose further coherent and convex model uncertainty measures. We take the coherent market risk measures introduced by Artzner et al. [2] as inspiration for new coherent model uncertainty measures. Convex model uncertainty measures are motivated by the reality of imperfect calibration in most practical applications. Branger & Schlag [20] have also used a Bayesian approach to measure model uncertainty. However, they use model-integration to aggregate market risk and model uncertainty into one measure by creating a ‘meta probability measure’ on the space of all scenarios and model parameters. This approach fails to differentiate between two very different types of risk — market and model — as we explain in Section 3.1.1.

The next subsection clarifies the distinction between (market) risk and (model) uncertainty. In the following two subsections we explore the possible sources of model uncertainty and review some of the existing literature on the topic. In Section 3.1.4 we review the financial setting for model uncertainty measures and in the subsequent section revise the Bayesian theory for computing posterior distributions in the context of model uncertainty measures. In Section 3.2 the concept of coherent model uncertainty measures is introduced. A simple example is used to show the shortcoming of Cont's worst-case measure [30] and then Bayesian posterior distributions are used to generalise this measure and construct new ones. In Section 3.3 the work is extended to convex model uncertainty

measures. A numerical example is worked through to compare the uncertainty values for different financial derivatives and the robustness of these values in Section 3.4. Finally Section 3.5 recommends possible extensions.

3.1.1 Risk vs Uncertainty

Although most authors refer to the cost of selecting the wrong model as model *risk* we deliberately refer to it as model *uncertainty*. This follows the reasoning presented by Cont [30]. Recall the setup presented in Section 2.1.1: the underlying asset S evolves on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is the set of future scenarios, \mathcal{F} is a filtration of Ω and \mathbb{P} is a probability measure on \mathcal{F} . Then, precisely speaking, *risk* corresponds to not knowing which future scenario $\omega \in \Omega$ will be realised, whereas *uncertainty* corresponds to not knowing the probability measure \mathbb{P} on \mathcal{F} .

This distinction between risk and uncertainty was made clear as far back as 1921 by Knight [97]. Although subtle, the difference between not knowing the future state and not knowing the probability of the possible future states is significant and investors will have aversions to both. In his heavily cited thesis, Ellsberg [57] studied the impact of aversion to uncertainty and used it to explain some of the anomalies of agents' decisions.

In this thesis, the model θ corresponds to the probability measure \mathbb{P} . For example, different local volatility functions $\sigma(S, t)$ attach different probabilities to possible evolutions of the underlying S . Thus, for example, the finding of Chapter 2 that a variety of very different local volatility models $\sigma(S, t)$ can be calibrated to the same market data, demonstrates how large (model) uncertainty can be. Typically it is very difficult in practice to assign probabilities to future outcomes, which in our case are the future path evolutions of S .

Using the technical definitions of risk and uncertainty, we can also introduce the concepts of option price risk and option price uncertainty. As Chapter 2 showed, model uncertainty can manifest as option price risk — lack of knowledge of what is the correct model (i.e. probability measure on Ω) means we are unsure what is the correct price of a financial option. On the other hand, option price uncertainty is the result of not knowing the probabilities assigned to the possible models. The rest of this chapter deals with trying to measure model uncertainty (or equivalently, option price risk) by eliminating option price uncertainty via assuming a measure — the Bayesian posterior — on the set Θ of models θ .

3.1.2 Sources of Model Uncertainty

As mentioned, model uncertainty literature is in its infancy and only began to receive attention about 15 years ago when Derman [44] published his research notes on model risk. In his paper Derman identifies seven distinct sources of model risk. In this section we will use the term ‘model risk’ rather than ‘model uncertainty’ when the paper cited does so.

The chief of these seven model risks is the potential inapplicability of modelling: we might not be able to use complex mathematical models to project stock movements, that brinkmanship or psychology might play a more important role — although models are evolving for this and behavioural finance is gathering momentum. [44] argues that a clear understanding of the situation is needed before the language of mathematics can be used to represent it.

A more typical source of risk is that an incorrect model is chosen. For example, some factors might have been forgotten. More typically, factors can be incorrectly modelled, e.g. as deterministic when they are stochastic or vice versa. Relationships between variables may be misconstrued. For example, correlation between driving processes or credit spreads may be mis-specified. Moreover, the model for a stock process in one market may not be appropriate for another stock process in a different market with differing levels of interest rates or volatility. Or a model in a stable market might become inappropriate in a time of financial crisis. Furthermore, when market frictions such as transaction costs and illiquidity are factored in, a model might no longer be applicable. For example, the Black Scholes model gives a price in frictionless markets with constant interest rates but in reality neither of these assumptions hold.

A third possible risk is that the model is correct but the analytical solution found is wrong. And a fourth risk [44] states is that, even if the correct analytical solutions is found, the numerical approximation may not be accurate enough. For example, some numerical schemes may have natural limits to the accuracy of their approximation. Another related source of risk is possible bugs in the software or hardware used by agents to implement the numerical schemes.

A more important risk is that the correct model is inappropriately used. As an example [44] notes that a Monte Carlo pricer may give the correct price but if used with too few simulations will not converge to the correct value. The user needs to monitor the variance of the tests.

The final source of risk [44] points out is the sensitivity of the model to the input data e.g. calibration prices. Often overlooked, robustness of the model — whether small changes in the assumptions of the model or observed data causes disproportionate changes in the prices or Greeks or any other model inference — is an important consideration. If market conditions change suddenly or the input data is actually very noisy, then it is imperative the model is robust enough to account for this.

Figlewski & Green [77] categorise sources of model risk into three distinct groups: model mis-specification, unobservability of input parameters and market violation of assumptions. Model mis-specification covers what Derman refers to as incorrect model choice. Unobservability of input parameters refers, for example, to not being able to see the volatility to input into the Black-Scholes formula or not being able to observe the jump density in a general Lévy process. This risk is sometimes referred to as model mis-estimation. For market violation of assumptions, [77] uses the example of arbitrage-free fair option values. [77] points out that most financial models find a fair price by constructing a continuous perfect replication strategy and setting the price of the option

to the cost of this strategy. However, markets are not always open and transaction costs mean that the cost of continuous rebalancing would be infinite.

Branger & Schlag [20] spend some time clarifying the difference between model risk and market incompleteness. They argue that, in an incomplete market the true data generating process may be known, but not all contingent claims are attainable so the equivalent martingale measure is not unique. However, under model risk, we do not even know the true data generating process. Conversely, model risk does not imply market incompleteness. For example, if model risk arises from several candidate deterministic volatility models, then each of these models is still complete but we still do not know which gives the perfect hedge.

3.1.3 Related Research on Model Uncertainty

We have already introduced the paper by Figlewski & Green [77] in the previous section. Here we review their empirical study into the market and model risk exposures faced by an agent trading European calls and puts. [77] considers different volatility forecasting methods based on historical data and applies the methods to four underlyings: S&P 500 index, 3 month US\$ LIBOR, 10-Year Treasury Yield, Deutschemark Exchange Rate. Their first finding is that the strategy of writing and holding option positions without hedging produces very large risk exposures, even over long horizons, and diversification does not significantly reduce this risk exposure. After daily delta rehedging was added to the portfolios, Figlewski & Green found that the standard deviation, mean, and worst-case returns were all reduced. However, worst case losses were still several times the initial premium, particularly for out-of-the-money contracts. They conclude that writing options with volatility markups (of up to 50%) turns a very risky trading strategy into a profitable one. By writing an option with a volatility markup we mean that the value of the volatility used in the calculation of the price is greater, i.e. ‘marked up’, than the volatility actually estimated from the data. This gives a price greater than would have been found with the original estimated volatility and is thus a safer price for an agent to sell the option for. This finding, [77] concludes, indicates that the model risk from mis-estimating volatility in trading and hedging derivatives positions is very large.

In contrast, Hull & Suo [86] look at the model risk from mis-specification of the model rather than mis-estimation. They consider the pricing errors arising from a continually recalibrated local volatility model. They price a compound option, a European call option on a European call option, and a barrier option. They find that the continually recalibrated local volatility model always correctly prices European style options, where the payoff is contingent on the asset price at just one time. However, for exotic options dependent on the distribution of the asset price at two or more times the model can perform badly.

Hull & Suo [86] argue that this failure of the local volatility surface is to be expected. They explain that the local volatility model is designed to match European options correctly but not options dependent on the value of the underlying asset at multiple times. Let $\phi_n(t_1, \dots, t_n)$ be the joint probability distribution of the asset price at times t_1, \dots, t_n and $\phi_1(t_1), \dots, \phi_n(t_n)$ the marginal distributions of the asset price at times t_1, \dots, t_n respectively. Then [86] points out that the local volatility model is designed so that all the marginals $\phi_1(t_1), \dots, \phi_n(t_n)$ are correct but in no way correctly reproduces $\phi_n(t_1, \dots, t_n)$ or any other joint probability distribution. And this is fundamental to why different local volatility surfaces can be fitted to the same calibration prices (marginals) as we saw in the previous chapter, but why these surfaces give very different prices for exotic and path dependent options (joint distributions). This point is further clarified by Britten-Jones & Neuberger [21] who show how very different volatility processes can be adjusted to fit the same observed option prices exactly — hence the prevalence of high model uncertainty.

Kerkhof et al. [96] look to quantify model risk with a view to determining how much regulatory capital should be set aside. Following the work of the aforementioned papers, they find that model risk due to mis-specification is much larger than the model risk due to estimation error. They specify model risk ρ as the difference between the worst-case market risk measurement method (RMM) and some reference market risk measurement method corresponding to reference model $m \in \mathcal{M}$:

$$\rho_{RMM}(\Pi, m, \mathcal{M}) = \sup_{k \in \mathcal{M}} RMM(\Pi_k) - RMM(\Pi_m), \quad (3.1)$$

where each model k in \mathcal{M} gives a different market risk $RMM(\Pi_k)$ for portfolio Π .

For the actual form of RMM , [96] suggests a number of alternatives: VaR, the coherent measures introduced by Artzner et al. [2], worst conditional expectation, tail conditional expectation. They then prove that, if RMM is translationally invariant, so too is ρ_{RMM} , and if RMM satisfies positive homogeneity, then so too does ρ_{RMM} . However, the same cannot be said for the property of subadditivity or monotonicity, which is why the authors cannot represent their worst-case measure as a coherent measure. This is something we address in this chapter.

Kerkhof et al. [96] go further and decomposes the model risk measure ρ_{RMM} into a parametric part (mis-estimation risk) and non-parametric part (mis-specification risk). As we clarify in the next section, following Cont's approach [30], we see the distinction between the two types as arbitrary and unnecessary. A ‘meta-model space’ could be constructed in which different parameters correspond to different models and then the two types of risk coincide.

We also remark here that [96] uses a worst-case approach and rejects a Bayesian approach, a sum of differently weighted risks $RMM(\Pi_k)$, on grounds that the choice of prior is difficult and arbitrary. But in this thesis, we have shown in the previous chapter that in certain situations a formulation for the prior can very naturally and unavoidably arise from the necessity for regularisation of the calibration problem.

Another paper which addresses the issue of measuring model risk is the paper by Branger & Schlag [20] already cited in the previous section. They extend the worst-case approach mentioned above to Bayesian model risk measures which are closer in spirit to the philosophy of this chapter. They index their set of candidate models as M_i and denote the probability of the i^{th} by $p(M_i)$. They write P_i for the probability measure on the set of future scenarios (asset price evolutions) corresponding to model M_i and $\rho^{P_i}(X)$ for some market risk measure ρ of contract X under measure P_i . Then they define two different Bayesian methods of integrating market and model risk: model integration and risk integration.

In the first method, [20] defines the weighted market measure P and the consequent risk measure as follows:

$$\rho(X) = \rho^P(X) \quad \text{where} \quad P = \sum_i p(M_i)P_i.$$

Observe the symmetry in the above expression. Model risk and market risk are treated identically. For example $\mathbb{E}^P[X]$ can be viewed as a double sum (or double integral in the infinite model and scenario case) over the different models and scenarios.

For the second method, risk integration, [20] defines the weighted market risk measure (c.f. weighted market measure) by

$$\rho(X) = \sum_i p(M_i)\phi(\rho^{P_i}(X))$$

for some model risk aversion function ϕ . ϕ is increasing and taken as convex if the decision maker is model risk averse, linear if model risk neutral, and concave if model risk preferring. $\phi(X) = x^n$ for different $n > 1$ are proposed as possible convex functions. The authors rightly argue that it is necessary to assess model risk when pricing or hedging. However, Branger & Schlag measure market risk and model risk together, whereas an agent might find it useful to have a value for each separately. Hence in this chapter we look at measures for model uncertainty rather than at measures for the combined market and model risks.

Contreras & Satchell [35] use a Bayesian approach to construct confidence intervals for the Value-at-Risk (VaR) measure. They design priors for the mean μ and standard deviation σ of $VaR(X)$ for some claim X , and update these statistics using the observed data. However, because VaR is not subadditive or convex, it can lead to anomalous values for a portfolio of options [2]. For example we can easily find 2 options such that the VaR of the portfolio of 2 options is greater than the sum of the individual VaRs — which is nonsensical because it should not be riskier for one to hold two options in one account than two options in two different accounts. For this reason we only consider coherent and convex measures in this chapter.

3.1.4 Financial Setup

We reproduce the financial notation of the previous chapter but with some small modifications and simplifications to suit the emphasis of this chapter.

Suppose we have an asset price process $S = (S_t)_{t \geq 0}$. Let Θ be the set of different models for explaining the evolution of the observed price process S . Choose a model $\theta \in \Theta$ for S so that, for any time $t \geq 0$,

$$S_t = S(S_0, t, (Z_u)_{0 \leq u \leq t}; \theta)$$

where S_0 is the time 0 price and Z is some stochastic process(es). In what follows we will write $S(\theta)$ when we want to emphasise the dependence of S on the model θ .

To price claims on $S(\theta)$, we first need to find suitable models θ . By finding suitable $\theta \in \Theta$ we mean that we need to *calibrate* the model to observed data. We can either estimate θ via some statistical analysis of past values of $S(\theta)$ or we can find the θ implied by some benchmark market prices $V = \{V^{(i)} : i \in I\}$ of claims $C = \{C_i : i \in I\}$ written on S . As in the previous chapter we consider the latter in this chapter also.

In practice we do not observe the ‘exact prices’ V but only know that they lie in intervals called *bid-ask spreads*. The size of the bid-ask spread, $|V^{bid} - V^{ask}|$, of a claim is the difference between the price an agent is willing to pay for it and the price an agent is willing to accept for it. As a minimum requirement for no-arbitrage the model needs to be calibrated so that it reproduces the benchmark prices to within their respective bid-ask spreads. If $\{\mathbb{E}^\theta[C_i] : i \in I\}$ are the values of claims $\{C_i : i \in I\}$ calculated using model θ , then we try to find

$$\Theta(I) = \{\theta : \forall i \in I \quad \mathbb{E}^\theta[C_i] \in [V^{(i)bid}, V^{(i)ask}]\}.$$

When $|\Theta(I)| > 1$, i.e. more than one calibrated model exists, the agent is faced with the risk of selecting the wrong element $\theta \in \Theta(I)$ to use to price another claim X . This is what we refer to as *model uncertainty*.

It is worth remarking again here that sometimes model selection is decomposed into selecting the model type and the corresponding calibrated model parameter. Following the definitions of [95] the choice of calibrated parameter is referred to as the *model uncertainty of X due to mis-estimation* or *estimation error* of X . The choice of model type is called the *model uncertainty of X due to mis-specification* or *specification error* of X . The estimation error plus specification error gives the total model uncertainty of X . However, as mentioned in the previous section, we could strictly speaking treat models with different parameters as different model types and embed all possible models into a larger model space. Hence, for what follows we do not decompose models into type and parameter.

3.1.5 Bayesian Posteriors

Bayesian theory can be used to estimate the value of an unknown parameter. It provides a rigorous framework for combining prior information with observations to calculate likely values. Suppose we have some prior information for model θ (for example that it has a particular form, or its unknown parameters are positive), summarised by a *prior* density $p(\theta)$ for θ . And suppose we observe some noisy data $V = \{V^{(i)} : i \in I\}$ related to θ by

$$V^{(i)} = f^{(i)}(\theta^*) + e^{(i)}$$

for all $i \in I$, where θ^* is the true parameter value, e_i are independent noises for $i \in I$, I is an index set and $f^{(i)}(\cdot)$ are known functions of θ . In what follows, the function p will depend upon its argument. Then $p(V|\theta)$ is the probability of observing the data V given θ and is called the *likelihood* function. An application of Bayes' rule (see [70] for a good introduction) gives that the *posterior* density of θ is

$$p(\theta|V) = \frac{p(V|\theta)p(\theta)}{p(V)}, \quad (3.2)$$

where $p(V)$ is given by

$$p(V) = \int p(V|\theta)p(\theta)d\theta.$$

The Bayesian posterior $p(V|\theta)$ is a measure of the likeliness in any particular value of the parameter θ based on prior assumptions and observed data.

Example 3.1.1. Take the observations $V^{(i)} = \frac{1}{2}(V^{(i)bid} + V^{(i)ask})$, the functions $f^{(i)}(\theta) = \mathbb{E}^\theta[C_i]$, the noises $e^{(i)} \sim N(0, \frac{S_0^2}{10^8}\delta_i^2)$ and the weights $w_i = [\frac{1}{4}|V^{(i)bid} - V^{(i)ask}|^2]^{-1}$. This gives us our likelihood function. For the prior we use a Gaussian density $\exp\{-\frac{1}{2}\lambda\|\theta\|^2\}$, where $\|\cdot\|$ is some norm function of the finite-dimensional parameter θ which summarises model θ (the abuse of notation is deliberately done to simplify the notation). The posterior then becomes

$$p(\theta|V) = k \exp \left\{ -\frac{1}{2\delta^2} \left[\frac{10^8}{S_0^2} \sum_{i \in I} \frac{4|\mathbb{E}^\theta[C_i] - V^{(i)}|^2}{|V^{(i)bid} - V^{(i)ask}|^2} + \lambda\|\theta\|^2 \right] \right\},$$

where k is a normalising constant, $\delta^2 = \sum w_i \delta_i^2$ and λ is a pre-defined constant indicating how strongly we believe in our prior assumptions.

3.2 Coherent Model Uncertainty Measures

Recently, Cont [30] has suggested properties that a measure of model uncertainty should have and then given a worst-case measure that satisfies these properties. We introduce coherent measures and the worst-case measure in the next few subsections. We then present an example where the worst-case measure cannot differentiate between two claims that have very different model uncertainty profiles. To resolve this a new set of axioms is proposed and a representation of model uncertainty

measures satisfying these axioms is proved. Finally, examples of measures satisfying these axioms are constructed using the Bayesian posterior.

3.2.1 Coherent Market Risk Measures

To later construct an alternative class of coherent model uncertainty measures, coherent market risk measures are first reviewed [2].

Suppose we have a set Ω of scenarios ω . Let \mathcal{X} be a set of real valued functions X on Ω and \mathbb{P} a probability measure on Ω .

Definition 3.2.1. A *market risk measure* is a mapping from \mathcal{X} into \mathbb{R} . A *coherent* market risk measure ρ is a market risk measure that satisfies the following four axioms:

mkrk.T translational invariance: $\forall X \in \mathcal{X}$ and $\forall a \in \mathbb{R}$, $\rho(X + a) = \rho(X) - a$.

mkrk.M monotonicity: $\forall X, Y \in \mathcal{X}$ with $X \leq Y$, we have $\rho(X) \geq \rho(Y)$.

mkrk.S subadditivity: $\forall X, Y \in \mathcal{X}$, $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

mkrk.P positive homogeneity: $\forall \lambda \geq 0$ and $\forall X \in \mathcal{X}$, $\rho(\lambda X) = \lambda \rho(X)$.

Theorem 3.2.2. (*Representation Theorem* [2]). A market risk measure $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is coherent if and only if there exists a family \mathcal{P}_ρ of probability measures on the set Ω of scenarios such that

$$\rho(X) = \sup_{\mathbb{P} \in \mathcal{P}_\rho} \{\mathbb{E}^{\mathbb{P}}[-X]\}.$$

Example 3.2.3. Examples of coherent market risk measures are:

- i) worst-case $\rho_0(X) = \sup_{\omega \in \Omega} \{-X(\omega)\}$.
- ii) average value $\rho_1(X) = \mathbb{E}^{\mathbb{P}}[-X(\omega)]$.
- iii) expected shortfall $\rho_\beta(X) = \frac{1}{\beta} \int_{\Omega} -X(\omega) 1_{X \leq q_X(\beta)} \mathbb{P}[d\omega]$ for some $\beta \in (0, 1)$ and where the quantiles q are given by $\beta = \mathbb{P}[X \leq q_X(\beta)]$.

Remark 3.2.4. A typical value for β in Example 3.2.3 iii) is $\beta = 0.05$ i.e. find the average loss in the worst 5% of scenarios. $\beta = 0.01$ and $\beta = 0.10$ are also sometimes chosen. From the definitions in Example 3.2.3 it can be shown that $\rho_0 = \lim_{\beta \rightarrow 0} \rho_\beta$ and trivially $\rho_1 = \lim_{\beta \rightarrow 1} \rho_\beta$.

Remark 3.2.5. β can be viewed as a risk-aversion parameter: the smaller the value of β the greater an agent's aversion to risk.

3.2.2 Cont's Axioms

Recall that we have claims C_i , with corresponding observable bid-ask spreads $[V^{(i)bid}, V^{(i)ask}]$ for $i \in I$, that we use as a calibration set, and a set of models Θ . By an abuse of notation let θ also represent the risk-neutral probability measure for asset price process S corresponding to the model θ for S . Now assume that

$$\forall \theta \in \Theta, \quad \mathbb{E}^\theta[C_i] \in [V^{(i)bid}, V^{(i)ask}] \quad \forall i \in I, \quad (3.3)$$

i.e. all measures $\theta \in \Theta$ reproduce benchmark options to within their bid-ask spreads.

Let $\mathcal{X} = \{X : \forall \theta \in \Theta, \mathbb{E}^\theta[|X|] < \infty\}$ be the set of all contingent claims that have a well-defined price in every model. Define Φ to be the set of admissible trading strategies for which, $\forall \phi \in \Phi$, $G_t(\phi) = \int_0^t \phi_u dS_u$ is well defined and a θ -martingale bounded from below θ -a.s. for all $\theta \in \Theta$. For simplicity, we assume the risk-free rate of growth is zero, so there is no discounting.

Cont [30] defines a function $\mu : \mathcal{X} \rightarrow [0, \infty)$ to be a coherent model uncertainty measure if it satisfies (3.3) and the following four axioms:

cont.1 For benchmark options, the model uncertainty is no greater than the uncertainty of the market price:

$$\forall i \in I, \quad \mu(C_i) \leq |V^{(i)bid} - V^{(i)ask}|.$$

cont.2 Dynamic hedging with the underlying does not reduce model uncertainty, since the hedge is model dependent:

$$\forall \phi \in \Phi, \quad \mu \left(X + \int_0^T \phi_t dS_t \right) = \mu(X).$$

But if the value of a claim can be totally replicated in a model-free way using only the underlying, then the claim has zero model uncertainty:

$$\begin{aligned} \text{if } \exists x \in \mathbb{R}, \phi \in \Phi \text{ s.t. } \forall \theta \in \Theta, \quad X = x + \int_0^T \phi_t dS_t \quad \theta\text{-a.s. then} \\ \mu(X) = 0. \end{aligned}$$

cont.3 Diversification does not increase the model uncertainty of a portfolio:

$$\forall X_1, X_2 \in \mathcal{X}, \forall \lambda \in [0, 1], \quad \mu(\lambda X_1 + (1 - \lambda) X_2) \leq \lambda \mu(X_1) + (1 - \lambda) \mu(X_2).$$

cont.4 Static hedging of a claim with traded options is bounded by the sum of the model uncertainty of that claim and the uncertainty in the cost of replication:

$$\forall X \in \mathcal{X}, \forall a \in \mathbb{R}^d, \quad \mu \left(X + \sum_{i=1}^d a_i C_i \right) \leq \mu(X) + \sum_{i=1}^d |a_i| |V^{(i)bid} - V^{(i)ask}|.$$

3.2.3 Worst-Case Measure

Cont [30] proposes the function

$$\mu_0(X) = \sup_{\theta \in \Theta} \{\mathbb{E}^\theta[X]\} - \inf_{\theta \in \Theta} \{\mathbb{E}^\theta[X]\} \quad (3.4)$$

as a coherent measure of model uncertainty, i.e. it satisfies (3.3) and his four axioms cont.1,2,3,4 of the previous section. The measure finds the difference between the highest and lowest prices in Θ . It is called the ‘worst-case’ measure because it finds the largest difference amongst the collection of prices $\mathbb{E}^\theta[X]$ for contract X .

3.2.4 Motivating New Measures

We now motivate the study for a different set of axioms and coherent model uncertainty measures. First of all, the worst-case measure given in the previous section does not distinguish between prices that are more and less likely: the price corresponding to a model which reproduces the calibration set of prices very close to their mid bid-ask price is more likely than a model that does not. To see this we work through the following example.

Example 3.2.6. Let $f_K(\sigma)$ be the Black-Scholes value of a European call on an underlying with initial price S , interest rate zero, dividend rate zero, volatility σ , maturity 1 and strike K . The constant volatility σ is unknown but assumed to belong to the interval $\Sigma = [0.1, 0.15]$. This is akin to the uncertain volatility model studied by Lyons [102] and Avellaneda et al. [5] and mentioned in Section 3.1. The difference in the Bayesian setting is that we say we attach uniform prior probability to the values of $\sigma \in \Sigma$ rather than that we are using an uncertain volatility model.

Taking 50001 equally spaced points in Σ , the value

$$F_K(\sigma) = [f_K(\sigma) - f_K(0.10)]/[f_K(0.15) - f_K(0.10)]$$

is evaluated for $K = S$ and $K = S/2$ for different values of σ in Σ . The normalised call value $F_K(\sigma)$ can be viewed as a portfolio of European calls and cash. The values are recorded in the histograms of Figure 3.1. Also plotted are the vegas (derivatives of price with respect to the volatility) $dF_S/d\sigma$ and $dF_{S/2}/d\sigma$ for different values of σ . By construction, the option values all lie in $[0, 1]$ (and attain the bounds) so the worst-case coherent model uncertainty values corresponding to (3.4) are

$$\mu_0(F_S) = \mu_0(F_{S/2}) = 1 - 0 = 1.$$

However, Figure 3.1 clearly shows that the normalised lower-strike call price $F_{S/2}$ is much more likely to fall in the interval $[0.00, 0.10]$ than the interval $[0.10, 1.00]$, so intuitively has much lower model uncertainty than the normalised at-the-money call F_S . An alternative way to see this is to consider the vega of each option for different values of the volatility. The vega for F_S is almost constant, so the price is equally sensitive to volatility misquotes in the interval Σ . However, the

vega for $F_{S/2}$ varies significantly with σ and is close to zero in the first half of the interval Σ , so the price is insensitive to volatility misquotes in the first half of the interval Σ . Thus, selecting the wrong model (volatility) for the pricing of the at-the-money European call is more likely to give a larger mis-pricing than for the European call with lower strike.

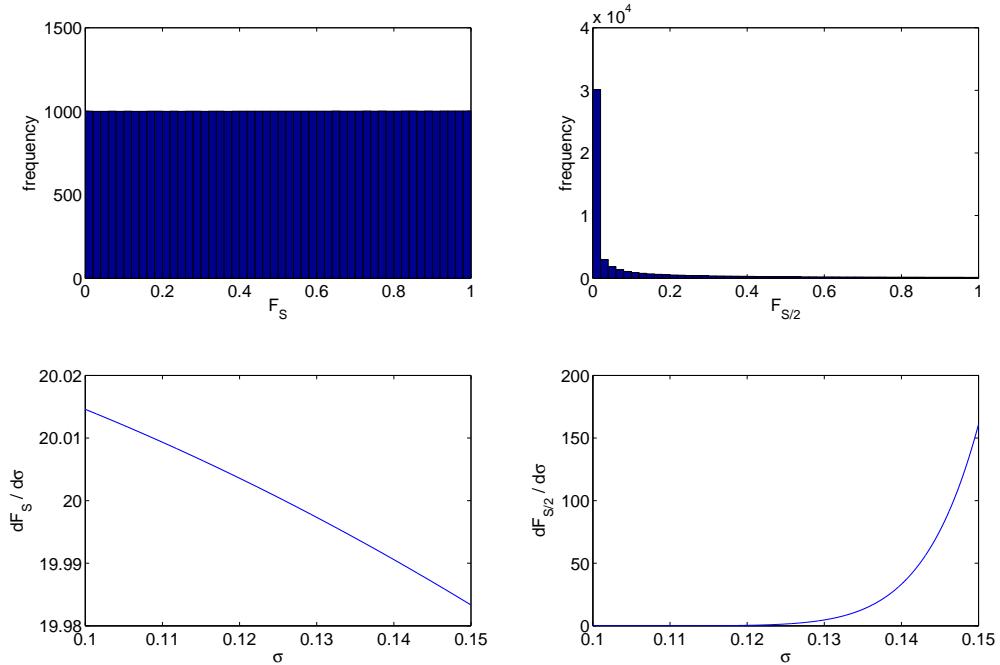


Figure 3.1: The upper two plots are histograms of prices for a normalised at-the-money European call and a normalised European call with lower strike. The volatility σ is assumed to be uniformly distributed on the interval $[0.10, 0.15]$. The lower two plots show the vega for each of the options.

The worst-case measure is insensitive to these subtleties, however, and cannot differentiate between the model uncertainty of the two claims. A measure which does account for such distributional disparities would be useful. Also, a method by which better information on the likelihood of parameter values can be taken into account would be helpful.

Secondly, although care has been taken to scale the model uncertainty measures defined by Cont's axioms cont.1,2,3,4 so that they are in real comparable monetary units, there is nothing to prevent the model uncertainty $\mu(X)$ of a claim X being greater than all possible discounted payoffs of the claim. We illustrate this by pricing a digital call option in the following example.

Example 3.2.7. Using the notation of the above example, suppose that there is only one calibration option — a one year at-the-money call option C . Recall the Black-Scholes pricing function $f_S(\sigma)$ for $\sigma \in \Sigma = [0.0, 0.5]$. We take the time 0 value of the underlying to be $S = 1$ and interest and dividend rates to be zero. Suppose that the quoted bid and ask values for C are $V^{bid} = f_S(0.0) = 0$

and $V^{ask} = 2f_S(0.5) = 0.3948$ respectively. Define the coherent model uncertainty measure

$$\mu'_0(X) = 2 \left(\sup_{\sigma \in \Sigma} \mathbb{E}^\theta[X] - \inf_{\sigma \in \Sigma} \mathbb{E}^\theta[X] \right).$$

Then μ'_0 still satisfies Cont's axioms cont.1,2,3,4. Consider now a one-year digital call option D with strike $1.1S$ and Black-Scholes pricing formula $d_{1.1S}(\sigma)$. Then the calculated model uncertainty value of D is

$$\begin{aligned} \mu'_0(D) &= 2(d_{1.1S}(0.5) - d_{1.1S}(0)) \\ &= 2(0.5237 - 0) \\ &> 1, \end{aligned}$$

which is nonsensical since the price of D must be less than or equal to 1, so the monetary value of the model uncertainty associated with the spread of prices for D must be less than or equal to 1.

This phenomenon is not restricted to the worst-case measure or multiples of the worst-case measure. The axioms of Cont, cont.1,2,3,4, are prone to scaling errors of the above example because there is no factor of scale in the axioms. Although axiom cont.1 ensures the model uncertainty values of calibration options do not exceed the size of their bid-ask spreads, none of the axioms guarantee that the model uncertainty value of another option is bounded above by the size of its bid-ask spread. And it is clear that we should sensibly expect the money value of the model uncertainty of a claim to be less than its maximum spread of prices. A set of axioms which includes this property would be seen as a useful amendment.

3.2.5 Alternative Axiomatic Approaches

We begin by observing that, by our definition of the contingent claim $X \in \mathcal{X}$, X should be thought of as a bivariate random variable $X = X(\theta, \omega)$ for $\theta \in \Theta$ and $\omega \in \Omega$. That is, the value of a claim X depends on the model θ it is valued in and the scenario ω . However, the price $\mathbb{E}^\theta[X]$ of claim X in model θ is given by

$$\mathbb{E}^\theta[X] = \int_{\Omega} X(\theta, \omega) \mathbb{Q}^\theta[d\omega],$$

where \mathbb{Q}^θ is the risk-neutral measure under model θ , and so it is the marginal of X (with respect to ω) and should be thought of as a univariate random variable on Θ . $\mathbb{E}^\theta[X]$ no longer depends on the realised scenario ω .

Recall from Section 3.1.1 that market risk and model uncertainty are two very different concepts. The first refers to lack of knowledge of the state, ω , and the second to lack of knowledge of the probabilities, θ , of each state. In this chapter we are interested in measuring model uncertainty only

and hence must integrate out the dependence of X on ω . This is precisely what we do by using the price map \mathbb{E}^θ :

$$\begin{aligned}\Theta \times \Omega &\xrightarrow{\mathbb{E}^\theta} \Theta, \\ X &\xrightarrow{\mathbb{E}^\theta} \mathbb{E}^\theta[X].\end{aligned}\tag{3.5}$$

$\mathbb{E}^\theta[X]$ is constant with respect to ω , and hence any measure of the variability of $\mathbb{E}^\theta[X]$ will be a measure of the price uncertainty of X and not the price risk of X . Effectively, we assume that once the model θ is fixed, the market risk is hedged away.

With this in mind, we then ask what exactly model uncertainty is trying to measure. It is clear that it is not the absolute values of $\mathbb{E}^\theta[X]$ that we are interested in but the *spread* of prices $\mathbb{E}^\theta[X]$. A number of approaches were considered before settling on the axioms defined in the next section. We mention a couple here and explain why they were rejected.

We considered measuring the distribution of $\mathbb{E}^\theta[X] - \inf_\theta \mathbb{E}^\theta[X]$. This was motivated by the worst-case measure which can be written as

$$\sup_{\theta \in \Theta} \left\{ \mathbb{E}^\theta[X] - \inf_{\theta' \in \Theta} \mathbb{E}^{\theta'}[X] \right\}.$$

The reasoning was that, instead of taking the supremum of $\mathbb{E}^\theta[X] - \inf_\theta \mathbb{E}^\theta[X]$, we could take the average, or expected shortfall or some other statistic with respect to a measure Q on Θ . This approach is akin to the model uncertainty measure (3.1) by Kerkhof et al. [96]. However, this approach ultimately failed because we were unable to assert a monotonicity axiom. There is no reason why $\mathbb{E}^\theta[X] - \inf_\theta \mathbb{E}^\theta[X] \leq \mathbb{E}^\theta[Y] - \inf_\theta \mathbb{E}^\theta[Y]$ should imply $\mu(X) \leq \mu(Y)$ (where μ is a model uncertainty measure), since $\inf_\theta \mathbb{E}^\theta[Y]$ might be very unlikely under measure Q , so using $\inf_\theta \mathbb{E}^\theta[Y]$ as a reference point is misleading.

Another possibility was to look at the distribution of $|\mathbb{E}^\theta[X] - E^{Q_X}[\mathbb{E}^\theta[X]]|$, so that the reference point is now a function of some measure Q_X on Θ . The logic was that the agent could choose the measure Q_X to match his risk preferences. For example, a very risk-averse agent might set $Q_X = Q_{\delta_{\theta'}}$, where $\theta' = \arg \inf_\theta \mathbb{E}^\theta[X]$ (which recovers the spreads of the approach in the previous paragraph), in order to capture the effect of the difference between the extreme prices. However, this approach was also rejected, because it involves the agent making two choices: the initial probability measure Q_X and the model uncertainty measure μ . Aside from the fact that this is duplication of choice, an agent might not be able to accurately assign a probability measure Q_X to her risk preferences.

3.2.6 New Axioms

Continuing the approaches of the previous section, we settle on the idea of finding the distribution of $|\mathbb{E}^\theta[X] - x|$ for some fixed point x to be described below. Note that we choose $|\cdot|$ rather than

$|\cdot|^2$ or $|\cdot|^{1/2}$, for example, because we wish the model uncertainty measure to have monetary units, not squared monetary units or square-root monetary units. If Q is a measure on Θ , then

$$E^Q[|\mathbb{E}^\theta[X] - E^Q[\mathbb{E}^\theta[X]]|]$$

is a measure of the absolute deviation of the prices with respect to the measure Q . With this in mind, we can list some of the properties we should expect our model uncertainty measures to have:

- function of spreads: by ‘spreads’ we mean the distribution of prices found by different models. We posit that the model uncertainty measure should be a function of the price spreads. For example, if every spread increases by 1 unit, then the model uncertainty should increase by 1 unit.
- monotonicity: if the spread of prices for X is less than the spread of prices for Y , then the model uncertainty of X should be less than that of Y .
- subadditivity: the spread of prices for a portfolio of two claims X and Y is not greater than the sum of the individual spreads, so neither should the corresponding model uncertainty values be.
- homogeneity: the spread of prices for X scales linearly with the number of claims X , but is indifferent as to whether the claim is bought or sold.

Moreover, since a zero spread of prices is preferable to a non-zero spread of prices, we should consider non-zero spreads unfavourable. To mathematically indicate this sentiment we explicitly consider functions of $-|\mathbb{E}^\theta[X] - x|$ instead of $|\mathbb{E}^\theta[X] - x|$. Furthermore, the modulus function $|\cdot|$ corresponds to a two-sided spread, i.e. $-|\mathbb{E}^\theta[X] - x|$ could correspond to $x + |\mathbb{E}^\theta[X] - x|$ or $x - |\mathbb{E}^\theta[X] - x|$. Hence, we will write the model uncertainty measure as a function of $-2|\mathbb{E}^\theta[X] - x|$ instead of $-|\mathbb{E}^\theta[X] - x|$ — to capture the fact that we are measuring a confidence interval rather than a one-sided distribution. To this end, define the set of possible spreads by

$$\hat{\mathcal{X}} = \{\hat{X}_x := -2|\mathbb{E}^\theta[X] - x| : X \in \mathcal{X}, x \in \mathbb{R}\}. \quad (3.6)$$

We formalise the above properties for coherent model uncertainty measures to describe a particular class of measures using the axioms given in the definition below.

Definition 3.2.8. A function $\mu : \mathcal{X} \rightarrow [0, \infty)$ is said to be a *coherent model uncertainty measure* with respect to a set of models Θ if $\mu(X) = \inf_x \{\hat{\mu}(-2|\mathbb{E}^\theta[X] - x|)\}$ for some function $\hat{\mu} : \hat{\mathcal{X}} \rightarrow \mathbb{R}$ satisfying the following four axioms:

mdun.T translational invariance: $\forall \hat{X} \in \hat{\mathcal{X}}$ and $a \in \mathbb{R}$, $\hat{\mu}(\hat{X} + a) = \hat{\mu}(\hat{X}) - a$.

mdun.M monotonicity: $\forall \hat{X}, \hat{Y} \in \hat{\mathcal{X}}$ with $\hat{X} \leq \hat{Y}$, we have $\hat{\mu}(\hat{X}) \geq \hat{\mu}(\hat{Y})$.

mdun.S subadditivity: $\forall \hat{X}, \hat{Y} \in \hat{\mathcal{X}}, \hat{\mu}(\hat{X} + \hat{Y}) \leq \hat{\mu}(\hat{X}) + \hat{\mu}(\hat{Y})$.

mdun.P positive homogeneity: $\forall \lambda \geq 0$ and $\forall \hat{X} \in \hat{\mathcal{X}}, \hat{\mu}(\lambda \hat{X}) = \lambda \hat{\mu}(\hat{X})$.

Remark 3.2.9. Note that implicit in this definition (by axiom mdun.T,P) is that the value of the model uncertainty for a claim is normalised so that its value is in monetary units and hence immediately comparable with the market price of the claim.

It was necessary to write the definition of μ as the infimum of the function $\hat{\mu}$ of spreads relative to fixed point x because, for example, choosing too large a value for x , say $x \rightarrow \infty$, sends all the values $\hat{X}_x = -2|\mathbb{E}^\theta[X] - x|$ to infinity for all bounded claims X and hence does not give a useful measure. Taking the infimum ensures that the comparisons between different claims X are consistent and meaningful. This approach is a generalisation of what was mentioned in the previous section except that the agent no longer chooses the value of x but it arises naturally through the optimisation of μ . And in fact the value of the optimiser x in Definition 3.2.8 turns out to be the Bayesian median price the agent with risk preferences corresponding to choice of measure μ should take for claim X .

Proposition 3.2.10. *If $\mu : \mathcal{X} \rightarrow [0, \infty)$ is a coherent model uncertainty measure (satisfies mdun.T,M,S,P) then μ also satisfies Cont's axioms cont.1,2,3,4.*

Proof. We prove that assumption (3.3) and axioms mdun.T,M,S,P imply Cont's axioms cont.1,2,3,4.

- cont.1: For benchmark option C_i , (3.3) implies $\mathbb{E}^\theta[C_i] \in [V^{(i)bid}, V^{(i)ask}]$. If $c = (V^{(i)bid} + V^{(i)ask})/2$, then $-2|\mathbb{E}^\theta[C_i] - c| \geq -|V^{(i)bid} - V^{(i)ask}| \forall \theta \in \Theta$, so by mdun.M

$$\begin{aligned}\mu(C_i) &= \inf_x \{\hat{\mu}(-2|\mathbb{E}^\theta[C_i] - x|)\} \\ &\leq \hat{\mu}(-2|\mathbb{E}^\theta[C_i] - c|) \\ &\leq \hat{\mu}(-|V^{(i)bid} - V^{(i)ask}|) \\ &= \hat{\mu}(0) + |V^{(i)bid} - V^{(i)ask}|,\end{aligned}$$

using mdun.T on the last line. Observe by mdun.P $\hat{\mu}(0) = 0$ so we recover cont.1.

- cont.2: follows because the measure is a function of $\mathbb{E}^\theta[X]$, i.e.

$$\begin{aligned}\mu\left(X + \int_0^T \phi_t dS_t\right) &= \inf_x \left\{ \hat{\mu}\left(-2\left|\mathbb{E}^\theta\left[X + \int_0^T \phi_t dS_t\right] - x\right|\right) \right\} \\ &= \inf_x \{\hat{\mu}(-2|\mathbb{E}^\theta[X] - x|)\} \\ &= \mu(X).\end{aligned}$$

Secondly, if $X = x_0 + \int_0^T \phi_t dS_t$, then $\mathbb{E}^\theta[X] = x_0$, where $x_0 \in \mathbb{R}$ so $\mu(X) = \inf_x \{\hat{\mu}(-2|\mathbb{E}^\theta[x_0 - x]|)\} = \hat{\mu}(0) = 0$, since the infimum is achieved at $x = x_0$.

- cont.3: immediate by mdun.S and mdun.P.

- cont.4: follows from mdun.S and mdun.H with cont.1.

□

3.2.7 Representation Theorem for Coherent Model Uncertainty Measures

We now prove a representation theorem, equivalent to the coherent market risk measures representation in Theorem 3.2.2, for coherent model uncertainty measures. Again assume the set of models Θ satisfies (3.3).

Theorem 3.2.11. *A function $\mu : \mathcal{X} \rightarrow [0, \infty)$ is a coherent model uncertainty measure if and only if there exists a family $\mathcal{Q}_\mu \subseteq \mathcal{Q}(\Theta)$ of probability measures on the set Θ of models such that*

$$\mu(X) = \inf_x \sup_{Q \in \mathcal{Q}_\mu} \{E^Q[2|\mathbb{E}^\theta[X] - x|]\}. \quad (3.7)$$

Proof. Recall $\hat{X}_x = -2|\mathbb{E}^\theta[X] - x|$ and $\hat{\mathcal{X}} = \{\hat{X}_x : X \in \mathcal{X}, x \in \mathbb{R}\}$. Then mdun.T,M,S,P correspond exactly to mkrk.T,M,S,P and hence we get the representation for $\hat{\mu}$ with respect to $\hat{\mathcal{X}}$ from Theorem 3.2.2 as

$$\hat{\mu}(\hat{X}_x) = \sup_{Q \in \mathcal{Q}_\mu} \{E^Q[-\hat{X}_x]\}.$$

And this implies the representation for μ to be

$$\mu(X) = \inf_x \sup_{Q \in \mathcal{Q}_\mu} \{E^Q[2|\mathbb{E}^\theta[X] - x|]\}.$$

□

Referring back to the previous two sections which motivated the development of new axioms, we see that the worst-case measure μ_0 satisfies the representation (3.7) for $\mathcal{Q}_{\mu_0} = \{Q_{\delta_\theta} : \theta \in \Theta\}$, where Q_{δ_θ} is an atomic measure on $\{\theta\}$ defined by Definition 2.2.2. However, the measure μ'_0 from Example 3.2.7 cannot admit a representation (3.7) for any family of measures \mathcal{Q} , because representation (3.7) is bounded by $\sup_\theta[X] - \inf_\theta[X]$. This is precisely what we concluded in Section 3.2.4 would be a desirable property for coherent model uncertainty measures. We prove the bound in the following corollary.

Corollary 3.2.12. *For all coherent model uncertainty measures μ and claims $X \in \mathcal{X}$*

$$0 \leq \mu(X) \leq \sup_{\theta \in \Theta} \mathbb{E}^\theta[X] - \inf_{\theta \in \Theta} \mathbb{E}^\theta[X].$$

Proof. First observe that $0 \leq 2|\mathbb{E}^\theta[X] - x|$ always, so by representation (3.7) $0 \leq \mu(X)$. Secondly, define $c = (\sup_\theta \mathbb{E}^\theta[X] + \inf_\theta \mathbb{E}^\theta[X])/2$; then $\forall Q \in \mathcal{Q}_\mu$

$$\begin{aligned}\inf_x \{E^Q[2|\mathbb{E}^\theta[X] - x|]\} &= \inf_x 2 \int |\mathbb{E}^\theta[X] - x| Q[d\theta] \\ &\leq 2 \int |\mathbb{E}^\theta[X] - c| Q[d\theta] \\ &\leq 2 \int \frac{1}{2} |\sup_\theta \mathbb{E}^\theta[X] - \inf_\theta \mathbb{E}^\theta[X]| Q[d\theta] \\ &= |\sup_\theta \mathbb{E}^\theta[X] - \inf_\theta \mathbb{E}^\theta[X]|.\end{aligned}$$

So by representation (3.7) we get the second required inequality. \square

3.2.8 Coherent Measures Using Bayesian Posteriors

With reference to Example 3.2.3, we define the *average value* model uncertainty measure (corresponding to ρ_1) as

$$\mu_1(X) = E^Q[2|\mathbb{E}^\theta[X] - M^{Q_1}[\mathbb{E}^\theta[X]]|],$$

where $M^{Q_1}[\mathbb{E}^\theta[X]]$ is the median value of $\mathbb{E}^\theta[X]$ with respect to probability measure on Q_1 on Θ . In the context of our Bayesian framework we associate Q_1 with the Bayesian posterior $p(\theta|V)$. μ_1 can be represented using Theorem 3.7 by taking

$$\mathcal{Q}_{\mu_1} = \{Q_1\},$$

i.e. the singleton Q_1 . We then see why the ‘ x ’ which minimises the spread for each X is given by $x = M^{Q_1}[\mathbb{E}^\theta[X]]$ — the Bayesian median price of the claim X with respect to measure Q_1 . It is because

$$\begin{aligned}0 &= \frac{d}{dx} E^{Q_1}[-2|\mathbb{E}^\theta[X] - x|] \\ &= \frac{d}{dx} \int_{\Theta} [-2|\mathbb{E}^\theta[X] - x|] Q_1[d\theta] \\ &= \int_{\mathbb{E}^\theta[X] > x} 2Q_1[d\theta] + \int_{\mathbb{E}^\theta[X] \leq x} -2Q_1[d\theta] \\ \Rightarrow \frac{1}{2} &= \int_{\mathbb{E}^\theta[X] > x} Q_1[d\theta] = \int_{\mathbb{E}^\theta[X] \leq x} Q_1[d\theta] \\ \Rightarrow x &= M^{Q_1}[\mathbb{E}^\theta[X]].\end{aligned}$$

Similarly, we can define the *expected shortfall* model uncertainty measure (corresponding to ρ_β) by

$$\mu_\beta(X) = \frac{1}{\beta} \int_{\Theta} 2|\mathbb{E}^\theta[X] - M^{Q_\beta}[\mathbb{E}^\theta[X]]| 1_{\mathbb{E}^\theta[X] \leq q_X(\beta/2) \vee \mathbb{E}^\theta[X] \geq q_X(1-\beta/2)} Q_1[d\theta];$$

here $M^{Q_\beta}[\mathbb{E}^\theta[X]]$ is the median value with respect to Q_β , the probability measure defined by $Q_\beta = \frac{1}{\beta} Q_1|_{\Theta_\beta}$, where $\theta \in \Theta_\beta$ if and only if $\mathbb{E}^\theta[X]$ belongs to the set

$$[\inf_{\theta \in \Theta} \mathbb{E}^\theta[X], q_X(\beta/2)] \cup [q_X(1 - \beta/2), \sup_{\theta \in \Theta} \mathbb{E}^\theta[X]]. \quad (3.8)$$

The quantiles q are given by $\beta/2 = Q_1[\mathbb{E}^\theta[X] \leq q_X(\beta/2)]$ and the probability measure Q_1 is again given by the Bayesian posterior $p(\theta|V)$. This time the measure admits the representation given by Theorem 3.7 by taking

$$\mathcal{Q}_{\mu_\beta} = \{Q' : Q' = \frac{1}{\beta} Q_1 1_{\Theta'} \text{ where } Q_1[\Theta'] = \beta\}, \quad (3.9)$$

and then we see that the ‘ $\inf_x \sup_{Q'}$ ’ occurs precisely for Q_β defined above and $x = M^{Q_\beta}[\mathbb{E}^\theta[X]]$ — the Bayesian median price of the claim X with respect to measure Q_β .

For completeness, observe the results, corresponding to Remark 3.2.4 for coherent market risk measures, that as $\beta \rightarrow 0$

$$\mu_\beta \rightarrow \mu_0, \quad (3.10)$$

$$\mathcal{Q}_{\mu_\beta} \rightarrow \mathcal{Q}_{\mu_0}, \quad (3.11)$$

and as $\beta \rightarrow 1$

$$\mu_\beta \rightarrow \mu_1, \quad (3.12)$$

$$\mathcal{Q}_{\mu_\beta} \rightarrow \mathcal{Q}_{\mu_1}. \quad (3.13)$$

Remark 3.2.13. In the coherent model uncertainty measures mentioned, the reference point ‘ x ’ has turned out to be the median $M^{Q_\beta}[\mathbb{E}^\theta[X]]$. It is unusual to talk about median prices and is more natural to look at mean prices, i.e. $E^{Q_\beta}[\mathbb{E}^\theta[X]]$. The reason for this unusualness is that we use the absolute difference $|\cdot|$ in $\hat{X}_x = -2|\mathbb{E}^\theta[X] - x|$ rather than the squared difference $|\cdot|^2$. This is because we are intent on the model uncertainty measure having units comparable to the price of the contract X , as we have stressed in Remark 3.2.9.

Example 3.2.14. Returning to the example we gave in Subsection 3.2.4, we can calculate the value of the model uncertainty for each of the two normalised call options under the new measures we have defined:

$$\begin{array}{lllll} \mu_0(F_S) = & 1.00 & \mu_{0.05}(F_S) = & 0.98 & \mu_{0.10}(F_S) = & 0.95 & \mu_1(F_S) = & 0.50 \\ \mu_0(F_{S/2}) = & 1.00 & \mu_{0.05}(F_{S/2}) = & 0.91 & \mu_{0.10}(F_{S/2}) = & 0.82 & \mu_1(F_{S/2}) = & 0.30. \end{array}$$

Both the average value and expected shortfall measures give significantly lower model uncertainty values for the option with strike $S/2$, which is what we would expect. They are able to differentiate between the two products and this is useful. Best practice would be a combination of these measures to make a judgement. Better still would be to build a profile of the model uncertainty measure by plotting $\mu_\beta(X)$ against $\beta \in [0, 1]$.

We have derived a whole range of different coherent model uncertainty measures and shown that they can be very informative. However, in some situations it is difficult to find models θ which satisfy (3.3), and we might be less strict about exactly matching the bid-ask prices. Under these circumstances, it is more typical to find models which reproduce calibration prices (reasonably) closely. For this reason we turn attention to convex model uncertainty measures.

3.3 Convex Model Uncertainty Measures

Cont's paper [30] extends coherent model uncertainty measures to convex model uncertainty measures, in a similar fashion that papers such as [69] and [66] extended coherent market risk measures to convex market risk measures. We first generalise Cont's definition of convex model uncertainty measures to the case where we observe bid-ask spreads and do not assume existence of a model that calibrates perfectly. We then state [30]'s calibration-error measure and generalise this measure using the Bayesian posterior.

3.3.1 Convex Market Risk Measures

The convex model uncertainty measure is motivated by the convex market risk measure [69]. We use the setup of Section 3.2.1, where recall Ω is a set of scenarios ω , \mathcal{X} is a set of real-valued functions X on Ω and \mathbb{P} is a probability measure on Ω .

Definition 3.3.1. A *convex* market risk measure is a market risk measure that satisfies the following three axioms:

mkrk.T translational invariance: $\forall X \in \mathcal{X}$ and $\forall a \in \mathbb{R}$, $\rho(X + a) = \rho(X) - a$.

mkrk.M monotonicity: $\forall X, Y \in \mathcal{X}$ with $X \leq Y$, we have $\rho(X) \geq \rho(Y)$.

mkrk.C convexity: $\forall X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$, $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$.

Recall the definition that the function ρ is lower semi-continuous if the set $\{X \in \mathcal{X} : \rho(X) \leq a\}$ is closed in \mathcal{X} for all $a \in \mathbb{R}$.

Theorem 3.3.2. (*Representation Theorem [69]*). $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is a finite valued and lower semi-continuous convex market risk measure if and only if there exists a convex penalty functional α and non-empty convex set \mathcal{P}_ρ of probability measures on the set Ω of scenarios such that

$$\rho(X) = \sup_{\mathbb{P} \in \mathcal{P}_\rho} \{\mathbb{E}^\mathbb{P}[-X] - \alpha(\mathbb{P})\}.$$

Remark 3.3.3. The special case when α only takes values 0 or ∞ corresponds to a coherent market risk measure.

3.3.2 Cont's Axioms

The motivation for convex measures is to penalise a measure according to some functional α . In the case of model uncertainty, instead of finding the set of models which are perfectly calibrated, i.e. satisfy (3.3), we consider all models but take them more or less seriously depending on how well they reproduce the observed market prices V .

In Cont's setup for convex model uncertainty measures he weakens assumption (3.3) by replacing it by the following version:

$$\exists \theta \in \Theta, \quad \mathbb{E}^\theta[C_i] \in [V^{(i)bid}, V^{(i)ask}] \quad \forall i \in I, \quad (3.14)$$

i.e. there exists a measure (model) θ that reproduces all benchmark options to within their bid-ask spreads. Cont [30] defines function $\mu^* : \mathcal{X} \rightarrow [0, \infty)$ by

$$\mu^*(X) = \sup_{\theta \in \Theta} \{\mathbb{E}^\theta[X] - \alpha(\theta)\} - \inf_{\theta \in \Theta} \{\mathbb{E}^\theta[X] + \alpha(\theta)\} \quad (3.15)$$

for α satisfying

$$\alpha(\theta) \geq \max\{V^{(i)bid} - \mathbb{E}^\theta[C_i], \mathbb{E}^\theta[C_i] - V^{(i)ask}, 0\} \quad (3.16)$$

for all $i \in I$ and models $\theta \in \Theta$. μ^* is then defined to be a convex model uncertainty measure if it satisfies the following axioms:

cont.1 For benchmark options, the model uncertainty is no greater than the uncertainty of the market price:

$$\forall i \in I, \quad \mu(C_i) \leq |V^{(i)bid} - V^{(i)ask}|.$$

cont.2 Dynamic hedging with the underlying does not reduce model uncertainty since the hedge is model dependent:

$$\forall \phi \in \Phi, \quad \mu\left(X + \int_0^T \phi_t dS_t\right) = \mu(X).$$

But if the value of a claim can be totally replicated in a model-free way using only the underlying, then the claim has zero model uncertainty:

$$\begin{aligned} & \text{if } \exists x \in \mathbb{R}, \phi \in \Phi \text{ s.t. } \forall \theta \in \Theta, \quad X = x + \int_0^T \phi_t dS_t \quad \theta\text{-a.s. then} \\ & \quad \mu(X) = 0. \end{aligned}$$

cont.3 Diversification does not increase the model uncertainty of a portfolio:

$$\forall X_1, X_2 \in \mathcal{X}, \forall \lambda \in [0, 1], \quad \mu(\lambda X_1 + (1 - \lambda) X_2) \leq \lambda \mu(X_1) + (1 - \lambda) \mu(X_2).$$

cont.4' Static hedging of a claim using long positions in traded options reduces the model uncertainty of the original claim:

$$\forall X \in \mathcal{X}, \forall a \in [0, 1]^{d+1} \text{ s.t. } \sum_{i=0}^d a_i = 1, \quad \mu\left(a_0 X + \sum_{i=1}^d a_i C_i\right) \leq \mu(X) + \sum_{i=1}^d a_i |V^{(i)bid} - V^{(i)ask}|.$$

The last axiom cont.4 had to be amended to cont.4' to restrict a to the set $\{a \in [0, 1]^{d+1} : a_0 + \sum_{i \in I} a_i = 1\}$, because a convex measure of risk cannot extrapolate the risk of a portfolio to a larger, proportional portfolio.

3.3.3 Calibration-Error Measure

In Cont's paper [30], the *calibration-error measure* μ_0^* is given by

$$\mu_0^*(X) = \sup_{\theta \in \Theta} \{\mathbb{E}^\theta[X] - \alpha_0(\theta)\} - \inf_{\theta \in \Theta} \{\mathbb{E}^\theta[X] + \alpha_0(\theta)\},$$

with the convex penalty functional α_0 defined by

$$\alpha_0(\theta) = \|\mathbb{E}^\theta[C] \triangleright V\|,$$

where $\|\cdot\|$ is a vector norm on $\mathbb{R}^{|I|}$ and

$$(\mathbb{E}^\theta[C] \triangleright V)_i = \max\{V^{(i)bid} - \mathbb{E}^\theta[C_i], \mathbb{E}^\theta[C_i] - V^{(i)ask}, 0\};$$

here we assume that there exists a model $\theta \in \Theta$ which reproduces all prices to within their bid-ask spreads (3.14), i.e. such that $\alpha_0(\theta) = 0$.

3.3.4 New Axioms

We would like to obtain a comparable representation to Theorem 3.2.11 for convex model uncertainty measures so that we may similarly expand our variety of convex measures. A larger variety of convex model uncertainty measures is likely to be far more useful to us since it is atypical to find a large set of perfectly calibrated models. In fact, it should not even be taken for granted that assumption (3.14) holds. So for our formulation we drop this assumption. As before, let $\hat{\mathcal{X}} = \{-2|\mathbb{E}^\theta[X] - x| : X \in \mathcal{X}, x \in \mathbb{R}\}$ be the set of spreads.

Definition 3.3.4. A function $\mu^* : \mathcal{X} \rightarrow [0, \infty)$ is said to be a *convex model uncertainty measure* with respect to a set of models Θ if $\mu^*(X) = \inf_x \{\hat{\mu}^*(-2|\mathbb{E}^\theta[X] - x|)\}$ for some lower semi-continuous function $\hat{\mu}^* : \hat{\mathcal{X}} \rightarrow \mathbb{R}$ satisfying the following three axioms:

mdun.T translational invariance: $\forall \hat{X} \in \hat{\mathcal{X}}$ and $a \in \mathbb{R}$, $\hat{\mu}^*(\hat{X} + a) = \hat{\mu}^*(\hat{X}) - a$.

mdun.M monotonicity: $\forall \hat{X}, \hat{Y} \in \hat{\mathcal{X}}$ with $\hat{X} \leq \hat{Y}$, we have $\hat{\mu}^*(\hat{X}) \geq \hat{\mu}^*(\hat{Y})$.

mdun.C convexity: $\forall \hat{X}, \hat{Y} \in \hat{\mathcal{X}}$ and $\lambda \in [0, 1]$, $\hat{\mu}^*(\lambda \hat{X} + (1 - \lambda) \hat{Y}) \leq \lambda \hat{\mu}^*(\hat{X}) + (1 - \lambda) \hat{\mu}^*(\hat{Y})$.

We chose to adjust the axioms provided by Cont [30] for convex model uncertainty measures for three reasons. Firstly, they are of simpler form than those given by Cont, insomuch as there are only three and they mirror the axioms for convex market risk measures. Secondly, using an identical proof to Corollary 3.2.12 we see that now we avoid the undesirable scenario of the convex model uncertainty of an option being greater than the spread of prices. For example, there is nothing in the axioms presented by Cont which prevents a digital call option from having convex model uncertainty measure greater than 1. And this is important because one of the objectives of the measure is to be correctly scaled to the monetary value of the option (see Remark 3.2.9). Thirdly, the axioms

given in Definition 3.3.4 allow the representation given by the theorem in the next section, and thus a more general construction. We can in fact recover Cont's axioms from the new ones presented in Definition 3.3.4.

Proposition 3.3.5. *If $\mu^* : \mathcal{X} \rightarrow [0, \infty)$ given by (3.15) is a convex model uncertainty measure (satisfies mdun.T,M,C), then μ^* also satisfies Cont's axioms cont.1,2,3,4'.*

Proof. We prove that assumption (3.14) with form (3.15) and axioms mdun.T,M,C imply axioms cont.1,2,3,4'.

- cont.1: observe by condition (3.16), for all $i \in I$ and $\theta \in \Theta$ we have

$$\sup_{\theta \in \Theta} \{\mathbb{E}^\theta[C_i] - \|V - \mathbb{E}^\theta[C]\|\} \leq V^{(i)ask} \quad \inf_{\theta \in \Theta} \{\mathbb{E}^\theta[C_i] + \|V - \mathbb{E}^\theta[C]\|\} \geq V^{(i)bid}.$$

Hence for each $i \in I$,

$$\begin{aligned} \mu^*(C_i) &= \sup_{\theta \in \Theta} \{\mathbb{E}^\theta[C_i] - \|V - \mathbb{E}^\theta[C]\|\} - \inf_{\theta \in \Theta} \{\mathbb{E}^\theta[C_i] + \|V - \mathbb{E}^\theta[C]\|\} \\ &\leq V^{(i)ask} - V^{(i)bid}, \end{aligned}$$

as required. Note also that, by assumption (3.14), $\exists \theta \in \Theta$ s.t. $\mathbb{E}^\theta[C_i] \in [V^{(i)bid}, V^{(i)ask}]$, and hence the model uncertainty $\mu^*(C_i)$ is positive.

- cont.2: as before, in the proof of Theorem 3.2.11 for the representation of coherent model uncertainty measures, since μ^* is a function of $\mathbb{E}^\theta[X]$, the first equality is true. The second equality follows from mdun.T again.
- cont.3: immediate by mdun.C.
- cont.4': follows from mdun.C and using $\mu^*(C_i) \leq |V^{(i)ask} - V^{(i)bid}|$ for all $i \in I$.

□

3.3.5 Representation Theorem for Convex Model Uncertainty Measures

Theorem 3.3.6. *A function $\mu^* : \mathcal{X} \rightarrow [0, \infty)$ is a convex model uncertainty measure if and only if there exists a convex penalty functional α and non-empty convex set $\mathcal{Q}_{\mu^*} \subseteq \mathcal{Q}(\Theta)$ of probability measures on the set Θ of models such that*

$$\mu^*(X) = \inf_x \sup_{Q \in \mathcal{Q}_{\mu^*}} \{E^Q[2|\mathbb{E}^\theta[X] - x|] - \alpha(Q)\}. \quad (3.17)$$

Proof. Recall $\hat{X}_x = -2|\mathbb{E}^\theta[X] - x|$ and $\hat{\mathcal{X}} = \{\hat{X}_x : X \in \mathcal{X}, x \in \mathbb{R}\}$. Then mdun.T,M,C for $\hat{\mu}^*$ finite-valued and lower semi-continuous correspond exactly to mkrk.T,M,C, and hence we get the representation for $\hat{\mu}^*$ with respect to $\hat{\mathcal{X}}$ from Theorem 3.3.2 as

$$\hat{\mu}^*(\hat{X}_x) = \sup_{Q \in \mathcal{Q}_{\mu^*}} \{E^Q[-\hat{X}_x] - \alpha(Q)\}.$$

for convex α and non-empty convex set Q . And this implies the representation for μ^* is

$$\mu^*(X) = \inf_x \sup_{Q \in \mathcal{Q}_{\mu^*}} \{E^Q[2|\mathbb{E}^\theta[X] - x|] - \alpha(Q)\}.$$

□

For example, for the calibration-error convex model uncertainty measure μ_0^* defined in Section 3.3.3, the corresponding convex set of measures is the convex set containing $\{Q_{\delta_\theta} : \theta \in \Theta\}$, i.e. $\mathcal{Q}_{\mu_0^*} = \mathcal{Q}(\Theta)$ — the set of all measures on Θ .

3.3.6 Constructing Penalty Functionals using Bayesian Posteriors

Although Cont's penalty functional α_0 is not arbitrary, we can use the rigorous Bayesian framework we set up in Section 3.1.5 to formalise this choice of functional and generalise it. This will give us a richer choice of convex model uncertainty measures for making decisions.

Example 3.3.7. Let $p(\theta|V)$ be as given in Example 3.1.1. Then μ_λ^* defined by

$$\begin{aligned} \mu_\lambda^*(X) &= \inf_x \sup_{Q \in \mathcal{Q}(\Theta)} \{2|\mathbb{E}^\theta[X] - x| - \alpha_\lambda(Q)\} \\ &= \sup_{\theta \in \Theta} \{\mathbb{E}^\theta[X] - \alpha_\lambda(\delta_\theta)\} - \inf_{\theta \in \Theta} \{\mathbb{E}^\theta[X] + \alpha_\lambda(\delta_\theta)\}, \end{aligned}$$

and the convex penalty functional

$$\begin{aligned} \alpha_\lambda(\delta_\theta) &= \left[-2\delta^2 \log \frac{p(\theta|V)}{k} \right]^{1/2} \\ &= \left[\frac{10^8}{S_0^2} \sum_{i \in I} w_i |\mathbb{E}^\theta[C_i] - V^{(i)}|^2 + \lambda \|\theta\|^2 \right]^{1/2}, \end{aligned} \quad (3.18)$$

for some weights w_i summing to 1, is a convex model uncertainty measure. Recall that we choose the 'regularisation' term $\lambda \|\theta\|^2$ precisely so that the square of expression (3.18) is convex in the unknown parameter θ .

Remark 3.3.8. If $\lambda = 0$, i.e. we use a totally uninformative Bayesian prior, then the penalty functional become $\alpha_0(\theta) = \|\mathbb{E}^\theta[C] - V\|$. This is akin to the calibration-error measure defined in Subsection 3.3.3.

Example 3.3.7 demonstrates how we can use a Bayesian posterior we have already computed to define a convex model uncertainty measure using Theorem 3.3.6. It is natural to use the Bayesian posterior to construct the penalty functional because the posterior is precisely a measure of our confidence in each model θ . In other words, if the posterior is greater, then our confidence in the model is greater, and so we should penalise the price corresponding to that model by less. We take the logarithm and re-scale the posterior so that it is in monetary units. The Bayesian posterior itself allows one to incorporate prior beliefs (such as smoothness or positivity) of the unknown model parameter θ via the prior $p(\theta)$, and is hence a useful extension of the theory introduced by Cont in [30].

3.4 Numerical Examples

We continue with the two local volatility examples of Chapter 2 provided by [89] and [28]. We price different options and numerically calculate the convex model uncertainty values.

Using the sample of surfaces plotted in Figure 2.1, we compute the convex model uncertainty values $\mu_\lambda^*(X)$ for five different claims X : three European call options with varying strikes (in-the-money, at-the-money, out-the-money), a barrier option (up-and-out call) and an American put option. The time 0 price of the underlying is 5000, and as before we take a constant interest rate of 0.05 and constant dividend rate of 0.03.

The values are calculated for $\lambda \in [0, 10^6]$ and plotted in Figure 3.2. The true prices (found using the true surface) are written in the legend of the graph. In the second graph, Figure 3.3, for confidence parameter $\lambda = 1$, the price found in each model for each option minus the true price is plotted in a histogram. Also drawn onto each histogram are the intervals associated with the ‘normalised deflated prices’ $\mathbb{E}^\theta[X] + \alpha_1(\theta) - \mathbb{E}^{\theta^*}[X]$ and ‘normalised inflated prices’ $\mathbb{E}^\theta[X] - \alpha_1(\theta) - \mathbb{E}^{\theta^*}[X]$.

We now repeat the first graph for the S&P 500 example given by Coleman et al. [28]. So, using the sample of surfaces plotted in Figure 2.6, we calculate the model uncertainty values $\mu_\lambda^*(X)$ for five similar claims X : three European call options of varying strikes (in-the-money, at-the-money, out-the-money), a barrier option (up-and-out put) and an American put option. Recall that the time 0 price of the underlying is \$590 and we take constant interest rate of 0.0600 and dividend rate 0.0262.

Some important observations can be made from these graphs. Firstly, for each option, as $\lambda \rightarrow \infty$, the model uncertainty $\mu_\lambda^*(X) \rightarrow 0$. This we should expect since $\lambda \rightarrow \infty$ implies that we are completely confident in our model and hence believe that no model error exists which $\mu_\lambda^*(X) \rightarrow 0$ as $\lambda \rightarrow \infty$ verifies.

A second interesting feature of Figure 3.2 is the *L-curve* likeness of the 5 option graphs. There is a noticeable kink in the curves where they go from being largely vertical to horizontal. This is less noticeable in Figure 3.4, which is a lot less smooth; there is however a noticeable defining kink at $\lambda = 10^{-0.5}$. As commented in Section 2.4.4 L-curves have been studied in the context of solving inverse problems using regularisation (see for example Hansen’s introductory treatment [80]) when plotting log-log graphs of smoothness of solution against fitting error of solution for different values of the regularisation parameter. Recall the discussion in Section 2.4.4 on using the *L-curve criterion* as a method for selecting the optimal regularisation parameter. The criterion proposes choosing regularisation parameter corresponding to the point of maximum curvature (‘corner’) on the L-curve. It would be interesting to investigate the implications of using such a criterion in the context of the convex model uncertainty graphs for the financial derivatives plotted.

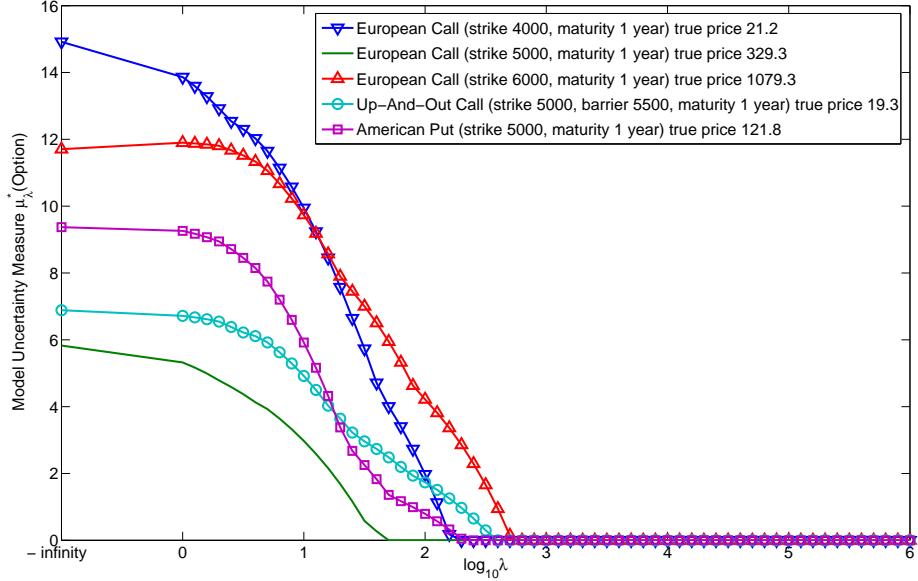


Figure 3.2: For simulated dataset: model uncertainty values for different options ($S_0 = 5000$) and varying values of the confidence parameter λ . The model uncertainty values are found using the set Θ corresponding to the sample of calibrated local volatility surfaces plotted in Figure 2.1.

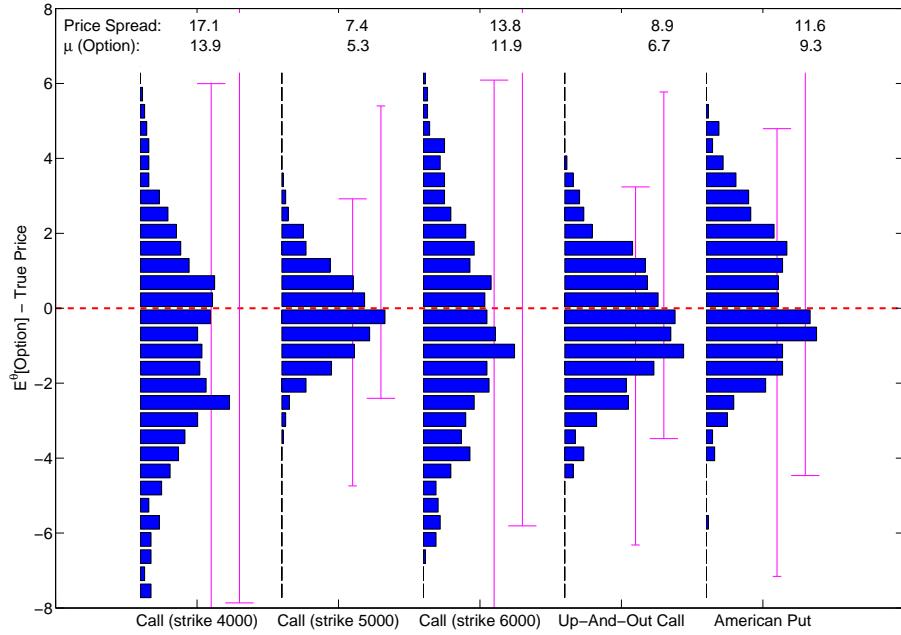


Figure 3.3: For simulated dataset: histograms of the difference between model prices and the true price. For each option the left line represents the range of values for the normalised deflated prices, the right line the normalised inflated prices. The supremum of the left line minus the infimum of the right line equals the model uncertainty value (for $\lambda = 1$).

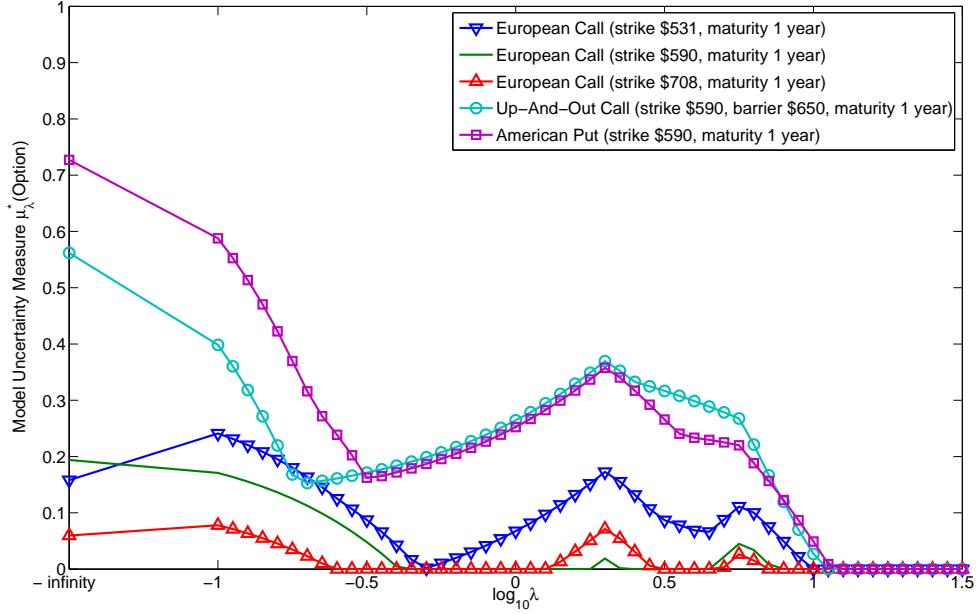


Figure 3.4: For S&P 500 dataset: model uncertainty values for different options ($S_0 = \$590$) and varying values of the confidence parameter λ . The model uncertainty values are found using the set Θ corresponding to the sample of calibrated local volatility surfaces plotted in Figure 2.6.

3.4.1 Robustness

Just as we did in Section 2.5.6 of the previous chapter, we shall repeat the model uncertainty calculations for the samples found using the different values of κ (form of the prior) and ε (observed data) used in Section 2.5.6. In Figure 3.5 and Figure 3.6, we plot the price spreads and model uncertainty values for each of the five options for varying κ and ε respectively.

Figure 3.5 shows that the model uncertainty values for the different options mostly stay the same for different values of κ . Especially as a method of comparing the model uncertainty of different contracts, we consistently see that the call with strike 4000 has the greatest model uncertainty value, followed by the call with strike 6000, then the American option, then the barrier option and finally the at-the-money call has the lowest model uncertainty. Only for $\kappa = 10^{-0.75}$ does the model uncertainty for the out-of-the-money call option become larger than for the in-the-money call option. This is likely to be a feature of the sampling but is worthy of future investigation.

Figure 3.6 implies a similar consistency between the ranking of different contracts. However there are two additional features. Firstly we note a downward trend in the model uncertainty values as the noise is increased. For example, for the call with strike 6000 the model uncertainty falls from approximately 13 for noise with mean $\varepsilon = 10^{-3}$ to approximately 9 for noise with mean $\varepsilon = 10^{-2.5}$. This is to be expected because, as we increase the noise in the observed prices, we are able to find

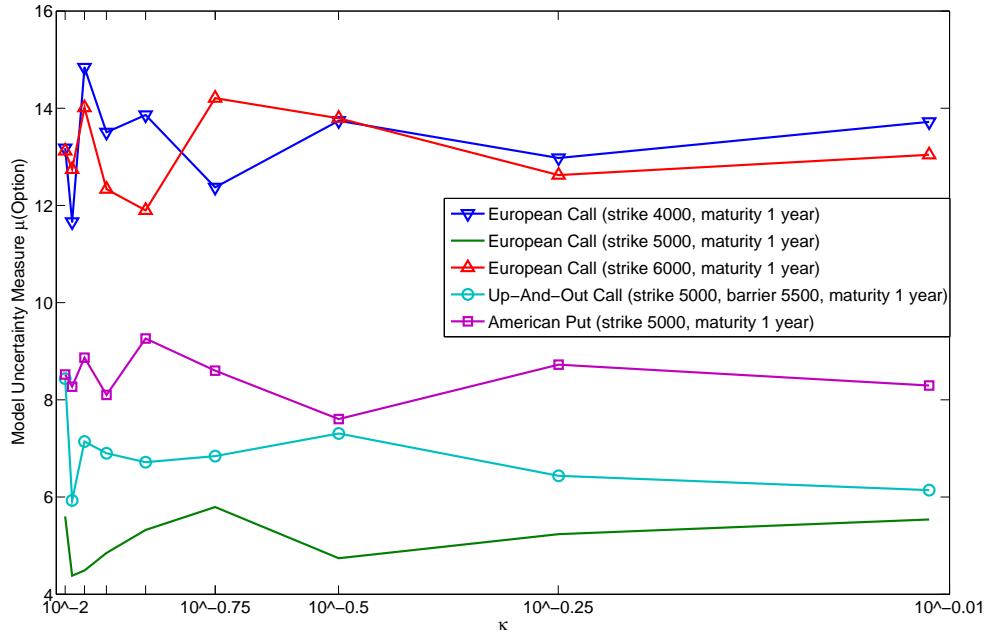


Figure 3.5: Model uncertainty values for 3 European call options, an American put option and an up-and-out barrier put option for different values of κ .

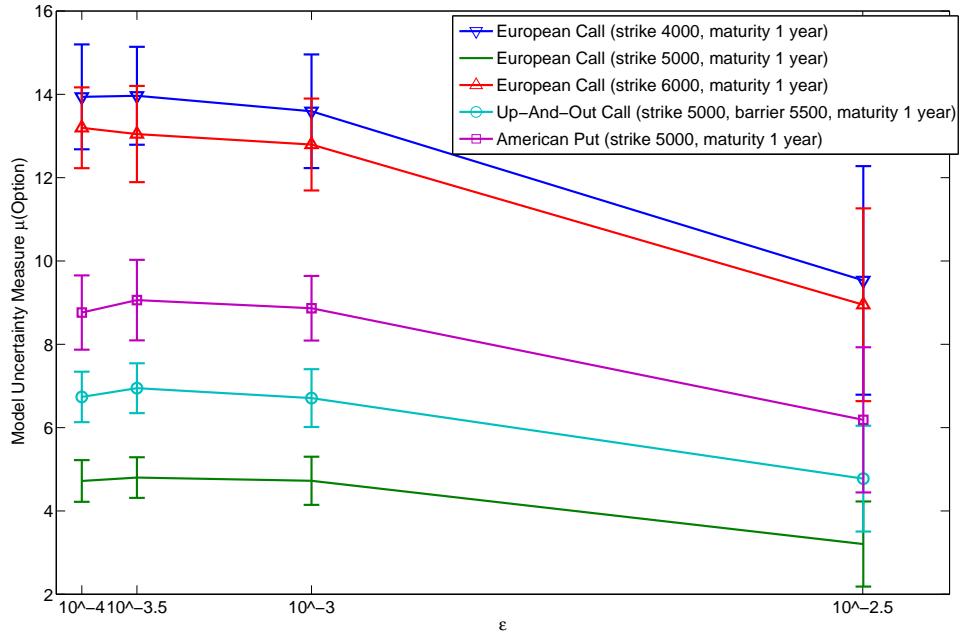


Figure 3.6: Model uncertainty values for 3 European call options, an American put option and an up-and-out barrier put option for different values of ε . For each ε the experiment was repeated 100 times with different noises so for each option we have included the mean model uncertainty values and the standard deviation error bars.

fewer calibrated surfaces and hence fewer models to price the options in, so less model uncertainty. A second feature is that the standard deviation of the model uncertainty values increases with the noise. Recall that for each level of noise we repeat the procedure for 100 different sets of noises. It is clear that for larger noises, different noises are going to cause much larger differences in the calibration prices, so in the distribution of calibrated models and consequent model uncertainty value.

In both cases, for deviations in the prior through adjusting κ and deviations in the observed data by adding noise with standard deviation ε , the model uncertainty values and ordering of the financial contracts are both fairly insensitive and robust.

3.5 Extensions

In this chapter we have extended the model uncertainty measures introduced by Cont [30] and generalised them to construct a wider variety of measures. We proved analogous representations for coherent and convex model uncertainty measures to those found by Artzner et al. [2] and Frittelli & Gianin [69] for market risk measures. Bayesian theory was then employed to assign posterior probabilities to candidate models and this distribution explicitly used to construct coherent and convex model uncertainty measures. Finally, numerical examples for the convex model uncertainty measures were provided and discussed.

It is worth remembering that, within the approach presented in this chapter, there are several parameters which can be changed to suit the preferences of the agent. Different prior beliefs can be incorporated by adjusting the prior density $p(\theta)$, calibration to different option types by changing the calibration set I of observed prices, greater confidence in prior beliefs by increasing the confidence parameter λ , to name only a few.

On the practical side of things it might be useful to design a ‘rule of thumb’ to aid the decision-making of practitioners. Some sort of binary ‘yes-no’ rule for judging whether a contract X has too much model uncertainty $\mu(X)$ or not would be a useful addition to this work. For example, the ‘Value at Risk’ (VaR) measure is often used to make the following kind of decision:

$$\text{reject } X \text{ if } \text{VaR}_{0.01}(X) > c_\rho,$$

i.e. do not invest in X if the losses of X in the worst 1% of scenarios are greater than c_ρ . So similarly we could construct an ‘acceptance-rejection’ rule based on model uncertainty measure μ such as

$$\text{reject } X \text{ if } \mu(X) > c_\mu,$$

i.e. do not invest in X if the model uncertainty $\mu(X)$ is greater than c_μ . For example, c_μ could be taken as 1% of the time-0 price V_0^X of X , depending on the investor’s risk preferences. Even better

would be to aggregate model risk and model uncertainty and have a joint rule such as

$$\text{reject } X \text{ if } \text{VaR}_{0.01}(X) + \mu(X) > c$$

for some c . Observe that if $\text{VaR}_{0.01}(X) + \mu(X) < 0$, then we would always invest because we make positive return in all combinations of market scenarios and models.

Something that has not so far been addressed is the dynamic aspect of the market. In the case of convex model uncertainty measures, as we observe different calibration prices each day, we have more information about the performance of each model to benchmark options, and hence by how much extra or less we should penalise each model; the penalty term thus becomes dynamic $\alpha \rightarrow \alpha(t)$. This would subsequently give us dynamic convex model uncertainty measures $\mu(t)$. Cont [30] raises this possibility but remarks that there is no guarantee of dynamic consistency, in the sense that the set of models calibrated to prices at time 0 may not be calibrated to prices at time t , unless updating includes adjusting the set of acceptable models i.e. $\Theta \rightarrow \Theta(t)$. Föllmer & Penner [65] have looked at time-consistent dynamic penalty functions in the context of market risk measures, so there is good reason to believe similar results can be achieved for model uncertainty measures. However, it is not yet clear what would be the corresponding axiomatic framework.

In this chapter we have only considered convex model uncertainty measures penalised by the magnitude of the misfit of the models to prior beliefs and the observed calibration prices — via the Bayesian posterior. However, it may be argued that the spread of costs of replication of a contract X in different models is a better indicator of model uncertainty. To this end we detail the use of Bayesian average hedging errors as penalty functions for convex model uncertainty measures in the next chapter, Chapter 4.

Chapter 4

Optimal Hedging

In this chapter we follow on from the work on calibration and model uncertainty measures of the two previous chapters. Using again the Bayesian posteriors built upon the calibration error, we design optimal hedges that minimise expected losses via different Bayesian loss functions designed according to an investor’s preferences. Comparisons made with standard hedging strategies show the Bayesian hedges to outperform these.

4.1 Introduction

Financial modelling and the study of financial markets is motivated by the desire to maximise profits and/or minimise risk. An agent is often more interested in reducing risk — the spread of possible profits — than she is in maximising the expected profit. For this reason, successful hedging has become a fundamental component of financial modelling. Hedging a position (in some financial instruments) is the process of negating the possible value fluctuations of the position caused by the realisation of different scenarios (in the financial markets). So, for example, a perfect hedging strategy causes the final value of the assets to be constant whatever scenario occurs. And for this reason (because it reproduces all payoffs of the original position), the initial value of a perfect (self-financing) hedging (replicating) strategy is used to determine the fair (no-arbitrage) price of the original position. Hence, the hedge gives the price, and corresponding hedging strategy, for a trader to use and thus is far more important than the price alone.

The practice of hedging is well studied in the financial mathematics literature and there are different types of strategies to negate different sources of risk. For example, the most common hedge, the Black-Scholes delta hedge [17], takes a position in the underlying stock to offset instantaneous first-order changes in the payoff of an option caused by changes in the value of the underlying. Similarly, a vega hedge takes a position in a similar (correlated) instrument to offset first-order changes in the payoff of the option caused by changes in the value of the volatility of the underlying. A delta-vega hedge would take positions in both the underlying and another instrument to offset first-order changes in the option value caused by both price risk and volatility risk.

In this chapter we look to optimise the simplest and most common hedge — the delta hedge — under model uncertainty. Using the distribution of different calibrated models found in Chapter 2 and observing the consequent model uncertainty shown in Chapter 3, we detail a method for finding the optimal hedging strategy. The different types of optimality are defined in Section 4.3, but essentially optimality can be in the sense of maximising the average hedging profit or minimising the average absolute hedging error or something similar. Because the practice in large financial institutions is to rehedge positions frequently, we consider dynamic hedging strategies rather than static ones. In the remainder of this introductory section we review some of the existing literature on hedging strategies. In the next section we revisit Bayesian theory from the point of view of hedging, and provide a motivating example. In Section 4.3 we detail the Bayesian loss functions associated with finding the optimal hedges and relate these loss functions to the convex model uncertainty measures of the previous chapters and also utility functions. Some computational details and numerical examples are given in Section 4.4. Finally, Section 4.5 recommends some extensions, including methods for customising the Bayesian loss functions to give the desired type of optimisation.

4.1.1 Delta Hedging using Deterministic Volatility Functions

Suppose we are trying to hedge a contingent claim X written on underlying S in a complete market using only the asset S and a risk-free asset B . By hedge we mean in particular that we are trying to construct a portfolio Π that replicates the payoff $h(S)$ of X , i.e. matches the value of $h(S)$ for all scenarios ω where the value of the claim depends on the scenario ω i.e. $S = S(\omega)$. Let f_t^X be the price of the claim X at time t . The replicating portfolio $\Pi = (\Delta, \Psi)$ consists of Δ units of S and Ψ units of B , where Δ and Ψ are chosen so that the value of Π after one time step δt matches the price $f_{t+\delta t}^X$ of X and the portfolio Π is self-financing. This is called *delta hedging*.

In their famous paper, Black & Scholes [17] model the underlying S as following a geometric Brownian motion with constant drift and volatility. They found the delta hedge to be equal to the sensitivity of the price of the claim X to the underlying S_t i.e.

$$\Delta_t = \frac{\partial f_t^X}{\partial S_t}.$$

The Black-Scholes model can easily be extended for the case where the volatility is a deterministic function of time t and/or asset S — what we call the local volatility function $\sigma(S, t)$ (which we detailed in Section 2.3). Under deterministic local volatility, S evolves according to

$$dS_t = \mu(S, t)S_t dt + \sigma(S_t, t)S_t dZ_t, \quad (4.1)$$

where Z is a Wiener process. It is apparent, however, that the underlying S does not usually behave as in the Black-Scholes model or extended deterministic volatility model, and many researchers have looked at the implications for this on the hedging-error distributions when using the Black-Scholes

delta. For example, El Karoui et al. [54] assume that the underlying follows a stochastic volatility model and price and hedge with a deterministic volatility function. They show that, for claims with convex payoffs, if the mis-specified deterministic volatility function dominates the true volatility, then so does the corresponding mis-specified price. Moreover, [54] finds that the hedging error in this case is always positive.

McIntyre [106] conducts a similar experiment but, instead of assuming a (stochastic volatility) process for the underlying, he uses real FTSE 100 index options data. McIntyre implements the Dupire formula (2.15) to find the deterministic local volatility function implied by option prices and computes the corresponding Dupire delta hedge ratios. A variety of 3-month at-the-money European call options are hedged over a 3-year period, using the Dupire deltas and the implied Black-Scholes deltas. [106] finds that the implied Black-Scholes deltas outperform the Dupire local volatility deltas. Furthermore, [106] observes that the Dupire approach is very unstable and sensitive to errors in the data, and consequently estimates of the derivatives of the option prices can be very erroneous and lead to greater hedging losses. Dumas et al. [50] reach a similar conclusion with their tests on S&P 500 options over a 5-year period. [50] fits the deterministic local volatility function using an implied trinomial tree approach.

On the other hand, the papers by Coleman et al. [27], [28] suggest that, following a careful fitting of the local volatility surface, the local volatility deltas do outperform the implied Black-Scholes volatility deltas. The papers calibrate a non-parametric local volatility surface to the prices by finding the values of the surface at a selection of ‘knots’ and then fitting a 2-dimensional natural cubic spline function through these knots. The authors find that the local volatility hedges easily outperform the implied Black-Scholes volatility hedges for simulated data, and is also better for hedging S&P 500 index and futures options. [27] argues that, especially for longer hedging horizons, the local volatility deltas are superior to the implied Black-Scholes deltas because implied deltas tend to be larger than the true hedge parameters.

Hull & Suo [86] conduct a numerical study of the effect of the implied deterministic local volatility function on hedging errors. The difference in [86] is that it considers what happens when the local volatility surface is recalibrated daily to observable market prices, as happens on trading desks in practice. [86] concludes that the hedging of vanilla options is less heavily dependent than American and exotic options on the model used. [86] remarks that the tests by authors such as McIntyre [106] and Dumas et al. [50] are not consistent with the recalibration procedure that actually takes place in practice, so it is no surprise that they find the static local volatility model performs badly in hedging. Like [106], [86] calibrates the local volatility function using Dupire’s formula, but unlike [106], it is recalibrated frequently. [106] fits a Heston stochastic volatility model to market data and uses this to price compound options (European call options on European call options) and barrier options. [106] finds that the recalibrated local volatility hedge performs well for the compound option but poorly

for the knock-out barrier option for some combinations of strike price and barrier level. This result is attributed by the authors to the barrier option's stronger path-dependence.

4.1.2 Hedging Volatility and Jump Risk

As remarked at the beginning of the previous section, it has been widely observed that asset prices do not usually evolve according to the local volatility model (4.1). Furthermore, it is certainly very unlikely that an asset price evolves to a time consistent local volatility function, as Hull & Suo's tests imply [86], since they concluded that a continuously recalibrated local volatility model hedges better than a local volatility model calibrated only once. Instead, it has been postulated that asset prices evolve according to non-deterministic volatility models possibly including jumps. As a consequence, there has been much research on finding ways to hedge the risk of a process having non-deterministic volatility with/without jumps. We review a selection of these papers in this section.

One well-studied non-deterministic volatility model, is the uncertain volatility model where the volatility σ_t at time t is known only to within an interval, i.e.

$$\sigma_{\min} \leq \sigma_t \leq \sigma_{\max}.$$

Two papers that concurrently solved the problem of hedging in this model are those by Lyons [102] and Avellaneda et al. [5]. [102] finds that the optimal risk-free trading strategy creates a ‘with profits’ policy where an unpredictable hedging profit is realised. This is akin to a super-replicating strategy. [5] reaches the same conclusion by solving a non-linear PDE called the Black-Scholes-Barenblatt equation, in which the delta-hedging volatility is chosen dynamically from the two extreme values σ_{\min} and σ_{\max} according to the convexity of the value function. The convexity is determined by the *gamma* which is the second derivative of the value function f_t^X with respect to the underlying price S_t . Their strategy is optimal in the sense that it is the least costly strategy that always returns a non-negative hedging profit. A third paper, by Avellaneda & Parás [6], tried to hedge the volatility risk by constructing a static hedge using an optimal combination of vanilla options and then dynamically delta-hedging the residual.

Using options to hedge the volatility risk is something that a number of other researchers, such as Hull & White [87] and Scott [123], have tried. Delta-vega hedging, in which a portfolio of underlying S , cash B , and vanilla option Y is constructed, is often used in stochastic volatility models to replicate derivative payoffs. We explain delta-vega hedging in more detail in Section 4.4.2 (see [116] or [43] for a complete introduction). However, something that has gathered momentum over the past two decade has been the use of volatility options to hedge volatility risk. Volatility options are contracts written on an implied volatility index (where it exists) or realised instantaneous volatilities, and have been extensively priced by Whaley [134], Grünbichler & Longstaff [78] and Detemple & Osakwe [47], for example. Recently, Psychoyios & Skiadopoulos [114] have conducted a numerical study into the

benefits of using volatility options to hedge. [114] simulates stochastic volatility paths and compares the delta-vega hedging performances using plain vanilla options and volatility options. The authors found that the use of volatility options as hedging instruments was not superior to the standard vanilla ones when hedging a European call option. However, [114] concedes that this might not be true when hedging exotic options.

As well as accounting for changes in the volatility of the underlying, many markets, such as equity markets and energy markets, exhibit jumps in the asset price. Carr & Wu [25] use a static hedge of five European call options for options on a simulated jump diffusion model and S&P 500 index options, and find that it outperforms daily delta hedging in all cases. More recently Cont et al. [34] and Tankov [129] have looked at the use of *quadratic hedging* for optimising hedging returns in the presence of jumps. Quadratic hedging is the strategy by which we minimise the variance of the hedging error. Hedge ratios are computed, for stocks and hedging instruments, which minimise the variance of the hedging error.

Föllmer & Leukert [63], [64] look at hedging in incomplete markets using *quantile hedging*. As with quadratic hedging, these papers minimise a statistic of the hedging error distribution. For an initial capital w_0 , the quantile hedge is the hedge which maximises the probability of a successful hedge (positive hedge error) under the constraint that the cost of the initial hedge is not greater than w_0 . Conversely, one can fix the probability p_0 and find the smallest initial capital that gives a successful hedge with probability p_0 . Partial hedging techniques are also considered by Jonsson & Sircar [93] in the setting of stochastic volatility.

4.2 Hedging in the Presence of Model Uncertainty

In this section we outline the problem of hedging in the presence of model uncertainty. In Chapter 2 we demonstrated how model uncertainty can arise, and in Chapter 3 we sought ways to quantify the model uncertainty for different contingent claims. In this chapter we use Bayesian techniques to find the optimal hedge when model uncertainty exists. First we review previous research on the effects of model uncertainty on hedging errors. We then formulate the problem of hedging when there is model uncertainty and motivate the potential pitfalls when using naive hedges. Finally we assess previous Bayesian solutions to the problem, and then present our Bayesian hedging strategy using loss functions.

4.2.1 Literature on the Impact of Model Uncertainty on Hedging

In the previous decade momentum has been gathering on the prevalence and effect of model uncertainty. A handful of papers have looked at the effect of model uncertainty on hedging errors. We remark here that there are two types of model uncertainty studied in the literature: model mis-specification (where the wrong model type is chosen) and model mis-estimation (where the wrong

model parameter is chosen). In the previous chapter we treated both types concurrently. However, in the context of this literature review it is more helpful to distinguish between the two.

We have already considered the first type of error, model mis-specification, in some detail in Section 4.1.1, where we reviewed research on the robustness of hedging strategies based on deterministic volatility models when the underlying evolves according to a stochastic volatility model or some other process. In this section we review some papers that actually try to quantify and understand the effect of this mis-specification rather than minimise it. Branger & Schlag [19] consider correcting the delta hedge for discretisation error and model specification error. They compute the expected hedging error and adjust the delta in order to achieve a mean-zero expected hedging error. However, this correction is only calculated for the case of Black-Scholes delta hedging when the underlying evolves according to a Heston stochastic volatility model. Psychoyios & Skiadopoulos [114] test using volatility options as hedging instruments and study the relationship between the complexity of the hedging model and the hedging performance. They find that the simplest model, which is also the most mis-specified model, provides the most reliable hedges. So, perhaps counterintuitively, increasing the complexity of the model does not necessarily reduce hedging errors.

An example of research into the implications of estimation error on hedging is Li's paper [101]. [101] considers the case of model calibration when exact calibrations do not exist. [101] finds analytically the sensitivity of the calibrated model parameters to the prices of the calibration instruments (in the same way agents calculate the Greeks — sensitivities of prices to market quantities). These sensitivities are used to set up a so-called 'instrumental hedge', whereby the derivatives portfolio is neutral to small changes in the prices of the hedging instruments. Changes in the prices of the hedging instruments cause changes in the calibrated model parameters and so this is similar to, for example, vega hedging, in which we hedge against changes in the volatility of the underlying. However, the difference is that even if the volatility does not change, because of market noise (bid-ask spreads), fluctuating liquidity, transaction costs, credit ratings, and so forth, the prices of the calibration instruments may still change and cause the model parameters to change. Hence, [101] argues that it is important to have an 'instrumentally hedged' portfolio. However, this approach is cumbersome because the parameter sensitivities are difficult to calculate and the cost of setting up such a large hedging portfolio is likely to offset the value of the hedging improvements. Furthermore, the philosophy of this thesis is not that the model changes, but that the model is unknown and hence that we should select the model whose corresponding hedging strategy optimises our hedging statistics.

Monoyios [108] has also looked at optimising hedging in the face of parameter uncertainty. [108] assesses the impact of uncertainty in the drift parameter in an incomplete market. It is shown that incorrect estimates of the drift parameter can lead to large hedging errors. [108] concludes that, depending on the sign of the mis-estimation of the drift parameter, the mis-estimation can

be benign or destructive. If the drift parameter is over-estimated, then the effect is small, but if the drift parameter is under-estimated then the effect can be catastrophic. [108] comments for a geometric Brownian motion process that, if the true drift is 20% per annum and the volatility is 20% per annum, then to get a 95% confidence interval for the drift parameter (from continuous observations of the process) takes approximately 1537 years! [108] proposes a filtering approach that incorporates learning in order to improve the performance of the hedging strategy. However, the solution necessitates solving a three-dimensional PDE, which is computationally difficult and makes testing over many simulated asset paths unfeasible.

4.2.2 Hedging Formulation

As in Chapter 2 suppose we observe a price process $S = (S_t)_{t \geq 0}$ and model it as a function of time t , some stochastic process(es) $Z = (Z_t)_{t \geq 0}$, and finite-dimensional parameter $\theta \in \Theta$, i.e.

$$S_t = S(t, \mathcal{F}_t, (Z_u)_{0 \leq u \leq t}; \theta), \quad (4.2)$$

or $S_t(\theta)$ for short and by abuse of notation of S . Let $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ be the filtration generated by Z , so S is an \mathcal{F} -adapted process.

Consider an option over a finite time horizon $[0, T]$ written on S and with payoff function h . Let the time- t value in the model θ of this option be written as $f_t(\theta)$, where we include the argument θ to emphasise the dependence of this price on the model parameter θ . As before we assume market completeness. Explicitly,

$$f_t(\theta) = \mathbb{E}[B^{-1}(t, T)h(S(\theta))|\mathcal{F}_t]$$

with respect to risk-neutral measure corresponding to θ , where $B^{-1}(t, T)$ is the discount factor for the time interval $[t, T]$, possibly stochastic.

Suppose at time $t \in [0, T]$ we observe a set of such option prices $\{V_t^{(i)}(\theta) : i \in I_t\}$, possibly with noise $\{e_t^{(i)} : i \in I_t\}$. In other words, we observe

$$V_t^{(i)} = f_t^{(i)}(\theta) + e_t^{(i)}, \quad (4.3)$$

for $i \in I_t$. Then, as discussed in Chapter 2, an associated calibration problem is to find the value of θ that most closely reproduces the observed prices $\{V_t^{(i)} : i \in I_t\}$, i.e. the parameter θ which minimises the error functional given in Chapter 2 as

$$G_t(\theta) = \frac{10^8}{S_t^2} \sum_{i \in I} w_i |f_t^{(i)}(\theta) - V_t^{(i)}|^2.$$

If now a calibrated parameter $\hat{\theta}$ has been found, one can begin to consider the prices of options with different payoffs and how best to hedge these options. For example, for a different claim X with payoff h^X on S it would be sensible to price it as

$$f_t^X(\hat{\theta}) = \mathbb{E}[B^{-1}(t, T)h^X(S(\hat{\theta}))|\mathcal{F}_t]. \quad (4.4)$$

Furthermore, taking a portfolio (Δ, Ψ) of stock S and cash B respectively to hedge the option, we have already commented in Section 4.1.1 that the standard Black-Scholes [17] delta at time t in our calibrated model would be given by

$$\Delta_t(\hat{\theta}) = \frac{\partial f_t^X(\hat{\theta})}{\partial S_t}. \quad (4.5)$$

However, as we have previously demonstrated, often we cannot know with certainty what is the correct calibration parameter $\hat{\theta}$. There is both too much noise in the prices and too few prices to calibrate the model. Add to this the fact that we do not even know the correct model type and we see that the choice for the delta is considerable.

Example 4.2.1. In the literature, there are two frequently cited delta hedges which, in the Bayesian framework with Gaussian observational noise we assume, can be referred to as the following:

1. $\Delta_t(\theta_t^{MLE})$ — the delta hedge corresponding to the *maximum likelihood estimator* (MLE) θ_t^{MLE} (see e.g. [86], [27], [106]), [50]). It minimises the calibration error, so is given by

$$\theta_t^{MLE} = \arg \min \{G_t(\theta)\}.$$

2. $\Delta_t(\theta_t^{MAP})$ — the delta hedge corresponding to the *maximum a posteriori* (MAP) estimator θ_t^{MAP} (see e.g. [89], [98], [37]). It minimises the regularised calibration error and is given by

$$\theta_t^{MAP} = \arg \min \{G_t(\theta) + \lambda \|\theta\|\}$$

for scalar confidence parameter λ and norm $\|\cdot\|$ described in Chapter 2.

Although not usually referred to as the ‘MLE’ and ‘MAP’ estimates, we showed in Remark 2.4.1 that minimising the error function and regularised error function is equivalent to maximising the likelihood and posterior densities respectively. The MAP estimate is more typically used in academia than the MLE estimate.

In this chapter we use the rigorous Bayesian framework in which we have already specified how likely a value of θ is based on prior knowledge of the parameter and the observed prices. In Chapter 2 we arrived at the situation whereby we had a distribution of prices but specified no hedge. In Section 4.2.5 we extend the Bayesian framework to calculate an optimal Bayesian hedge which minimises the expected hedging loss statistic — where expectation is taken both with respect to the different paths ω and candidate models θ .

Note that we solve the model selection problem by finding the optimal pricing-hedging program. In other words, we find the price (and accompanying hedging strategy) of an option X corresponding to the hedging strategy which minimises the expectation of a hedging loss statistic. A popular alternative for comparing hedging strategies is to fix the price V_0^X of the option and find the best hedging strategy starting with initial wealth V_0^X e.g. [55] and [86]. Proponents of this methodology

argue that charging a greater premium will naturally improve the distribution of terminal hedging errors. However, in this thesis we compare the hedging *returns* — the hedging error expressed as a percentage of the premium charged for the option. Proponents of this approach, such as [108], suggest that it is acceptable for the agent to incorporate other sources of risk into the premium the agent charges for option X . In our case, we choose to compare different pricing-hedging programs, rather than a straight-forward comparison of hedging strategies with equal initial wealths, because model uncertainty is an additional source of risk we wish to factor into the price of the options.

4.2.3 Motivating Examples

Imagine we are in a Black-Scholes [17] setting and observe the spread $[V^{bid}, V^{ask}] = [23.958, 24.103]$, for a 1-year European call option with strike 80 on underlying with time 0 price 100 and constant riskless rate of growth 5%. Unknown to us the true volatility is actually 0.15. Our prior belief of the volatility is given by the density in the left-hand plot of Figure 4.1: we believe volatility is small. By constructing a suitable posterior function, we find the corresponding MAP hedge and graph the histogram of the delta hedging profits (computed on 2000 true paths of the underlying) in the right-hand plot of Figure 4.1. Also shown is the performance of the Bayesian μ -hedge (which is explained later in Section 4.3).

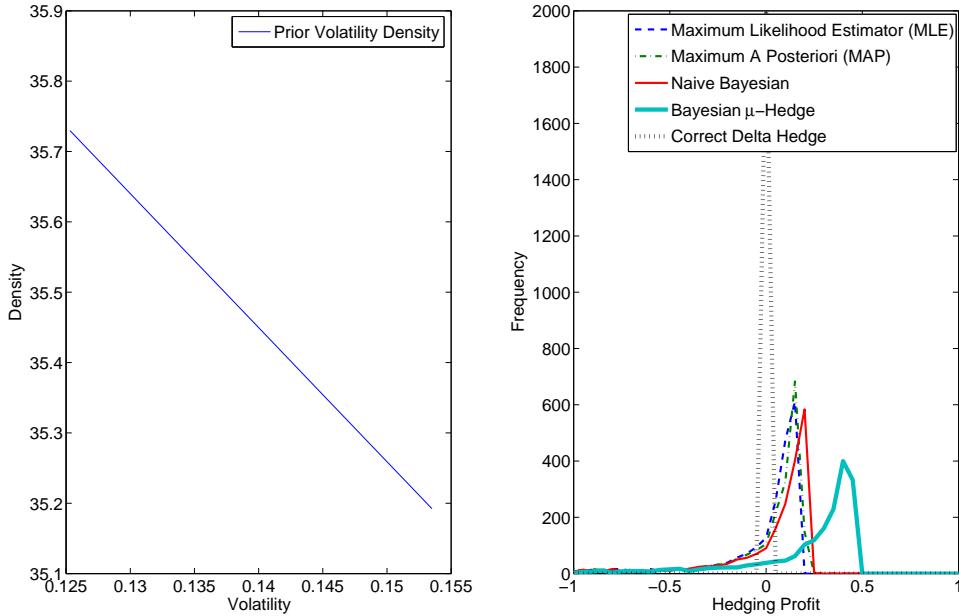


Figure 4.1: Density corresponding to prior view of volatility (left) and performance of different delta hedges (right) computed on 2000 true paths of the underlying. The true paths were evolved using the volatility value of 0.15.

As the right hand plot shows, the typical MAP hedge performs poorly compared with the

Bayesian μ -hedge. The Bayesian μ -hedge gives larger relative hedging profits than the MAP hedge, MLE hedge and naive Bayesian hedge (which is explained later in Section 4.2.5). To emphasise this point we repeat the experiment with all the same parameters except for a smaller true volatility value. This time the true volatility is 0.13. The bid-ask spread is the same as before, as is the prior density. Figure 4.2 shows that again the Bayesian μ -hedge outperforms the MAP, MLE and naive Bayesian hedges.

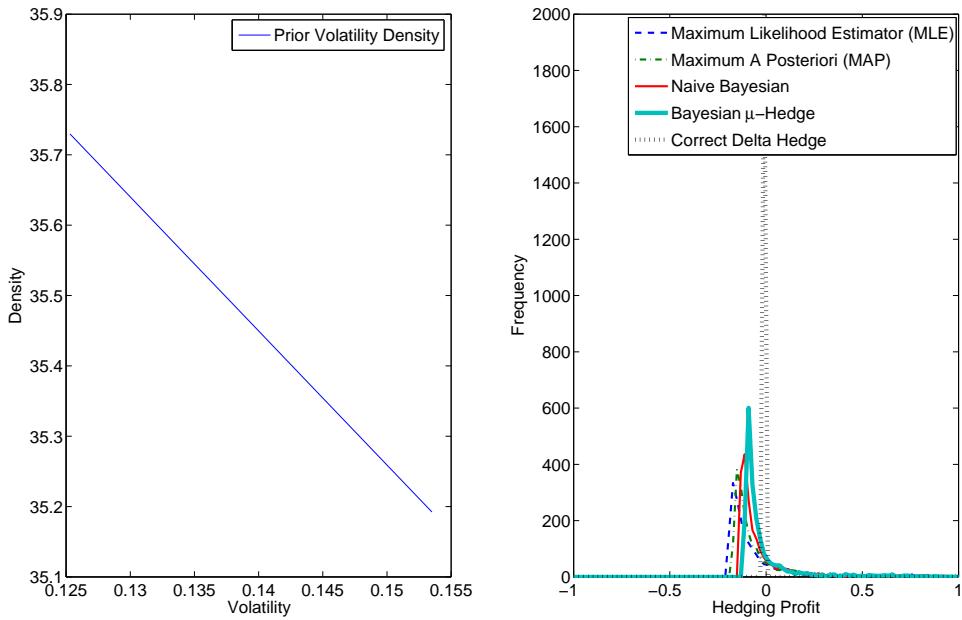


Figure 4.2: Density corresponding to prior view of volatility (left) and performance of different delta hedges (right) computed on 2000 true paths of the underlying. The true paths were evolved using the volatility value of 0.13.

4.2.4 Previous Bayesian Approaches to Hedging

As mentioned previously, Branger & Schlag [20] look at Bayesian ways to incorporate model uncertainty into their hedging strategies. One approach they specify is a simple Bayesian model average of the hedge ratios computed in different models. However, as we discuss in Remark 4.2.2 in the next section, there is no reason why this should give an admissible hedging strategy or be optimal.

[20] also suggests two Bayesian methods of aggregating market risk and model uncertainty for a claim X . In the first case they take risk measure $\rho(X) = \rho^{\mathbb{P}}$, where the large probability distribution \mathbb{P} is given by $\mathbb{P} = \sum_i p(M_i)\mathbb{P}_i$ with \mathbb{P}_i the market measure corresponding to model M_i and $p(M_i)$ the Bayesian prior probability of model M_i . In the second method they take the expectation of some function ϕ of the risk $\rho^{\mathbb{P}_i}$ with respect to the Bayesian prior on the set of candidate models, i.e.

$\rho(X) = \sum_i p(M_i)\phi(\rho^{\mathbb{P}_i}(X))$. They then choose the risk-minimising hedge for different risk measures $\rho^{\mathbb{P}_i}$, such as the expected shortfall and worst-case risk measures we detailed in Chapter 3.

However, in the view of this thesis, the aim should be to maximise the hedging profits — not to minimise the risk measure. Indeed, when hedging is concerned, the uncertainty measure should correspond to the distribution of hedging errors. Whereas in the previous chapter the model uncertainty measures were functions of price distributions, in this chapter we seek to extend this theory to functions of the distribution of hedging errors. We come back to this in Section 4.3.2 where we construct new convex model uncertainty measures by specifying penalty functionals corresponding to hedging errors.

Darsinos & Satchell [41] assume a Black-Scholes setting and use historical data to construct Gaussian and Inverted-Gamma-1 prior distributions for the drift and implied volatility parameter respectively. [41] uses 15 to 30 days worth of historical at-the-money European call prices. With these priors, analytical solutions for the posterior density for the Black-Scholes option price are found and used to forecast next-day option prices. As [41] remarks, this posterior can be combined with a loss or utility function to allow a decision to be made following the minimisation of the expected loss or maximisation of the expected utility. Although [41] does not study hedging explicitly, the recommended approach is similar to our method presented in the next section. However, we generalise from the Black-Scholes setting and present a generic strategy which can be applied to any model.

4.2.5 Bayesian Hedging using Loss Functions

Recall the setup of the previous two chapters. Suppose we wish to estimate the value of some finite-dimensional parameter θ . Assume we have some prior information for θ (for example that it belongs to a particular space, or is positive, or is a smooth function), summarised by a *prior* density $p(\theta)$ for θ . And suppose we observe some noisy data $V = \{V^{(i)} : i \in I\}$ related to θ by

$$V^{(i)} = f^{(i)}(\theta^*) + e^{(i)}$$

for all $i \in I$, where θ^* is the true parameter value, $e^{(i)}$ is some random noise, I is an index set and $f^{(i)}(\cdot)$ is a known function of θ . In what follows, by abuse of notation, the function p will depend upon its argument. Then $p(V|\theta)$ is the probability of observing the data V given θ and is called the *likelihood* function. An application of Bayes' rule gives the *posterior* density of θ as

$$p(\theta | V) = \frac{p(V|\theta)p(\theta)}{p(V)}, \quad (4.6)$$

where $p(V)$ is given by

$$p(V) = \int p(V|\theta)p(\theta)d\theta.$$

The Bayesian posterior is a measure of the confidence in any particular value of the parameter θ based on prior assumptions and observed data. For an explicit formulation, recall Example 3.1.1 from Chapter 3. Having found the posterior distribution $p(\theta|V)$, one can price and hedge other options using this distribution. First recall the following concepts of loss functions and Bayes estimators introduced in Chapter 2 Section 2.1.4.

The *loss function* $L(\theta, \theta')$ gives the loss incurred by taking θ' as the estimator for θ . It must satisfy

$$\begin{cases} L(\theta, \theta') = 0 & \text{if } \theta' = \theta, \\ L(\theta, \theta') > 0 & \text{if } \theta' \neq \theta. \end{cases}$$

Then given data V and loss function L , the corresponding *Bayes estimator* $\theta_L(V)$ is the value of θ which minimises the expected loss with respect to the posterior, i.e.

$$\theta_{L,p}(V) = \arg \min_{\theta'} \left\{ \int L(\theta, \theta') p(\theta|V) d\theta \right\}. \quad (4.7)$$

Suppose now that $L(\theta, \theta')$ corresponds to some measure of the hedging error caused by hedging contract X using parameter θ' when the correct hedge is found using parameter θ . Clearly we would like to find the value of $\hat{\theta}$ which minimises the (expected) hedging error. Using the definitions above, a sensible choice would therefore be to set

$$\hat{\theta} = \theta_{L,p}(V).$$

Then, for example, the Bayesian Black-Scholes hedge for option X at time t would be given by the Black-Scholes delta $\Delta_t(\hat{\theta})$ and the corresponding option price at time t would be $f_t^X(\hat{\theta})$, where Δ_t and f_t^X are given by (4.5) and (4.4) respectively.

Remark 4.2.2. It might at first seem intuitive to use Bayesian model averaging and take parameter

$$\bar{\theta} = \int \theta p(\theta|V) d\theta,$$

then hedge or price using this value, or directly take the delta hedge (Branger & Schlag [20]) to be

$$\bar{\Delta} = \int \Delta(\theta) p(\theta|V) d\theta.$$

However, this is a somewhat naive approach in the financial context, since a price should correspond to the best hedging strategy and there is no guarantee or intuition for why the above parameter or hedge would give the optimal hedging strategy. We cannot even be sure that $\bar{\theta}$ reproduces the observed data V or that $\bar{\Delta}$ corresponds to a calibrated parameter θ .

Although both the MAP and MLE estimators given in Example 4.2.1 can be recast as solutions of (4.7) for suitable choices of the prior and loss function respectively (see Section 2.1.4), in this chapter we wish to explicitly compare the performance of $\hat{\theta} = \theta_{L,p}(V)$ for a hedging error loss function L (to be detailed in Section 4.3) against the performance of θ^{MAP} .

4.3 Hedging Error Loss Functions

Consider a general (dynamic or static) hedging strategy given by a portfolio with time- t value Π_t . The portfolio is used to hedge an option X written on S with payoff $h^X(S)$ at maturity, time T . For simplicity, suppose that only S and the riskless asset B can be used to hedge. Then for some later time t the relative hedging error on any path realisation of S is given by

$$E_t = \frac{\Pi_t - f_t^X}{f_0^X},$$

where f_t^X is the calculated price of the option at time t . In particular, when the stock evolves according to one model, i.e. $S = S(\theta)$, and the hedging strategy is calculated using a different model, e.g. $\Delta = \Delta(\theta')$, then at time t the relative hedging error on any path realisation is given by

$$E_t(\theta, \theta') = \frac{\Pi_t(S(\theta), \Delta(\theta')) - f_t^X(\theta)}{f_0^X(\theta)}. \quad (4.8)$$

Since in the financial context, no-arbitrage arguments demand the price for an option to equal the initial value of a self-financing replicating portfolio (i.e. hedging strategy), it is natural to consider using hedging errors to construct Bayesian loss functions. To this end we take

$$L^g(\theta, \theta') = \mathbb{E}^\theta[g(E_T(\theta, \theta')) | \mathcal{F}_0] \quad (4.9)$$

as our loss function, for some function g of the random variable $E_T(\theta, \theta')$ (random with respect to the set Ω of scenarios ω) representing the relative terminal hedging error. Note that the expectation is taken in the risk-neutral measure corresponding to model θ .

Example 4.3.1. Different choices of g give common hedging performance indicators:

1. $g_\mu(z) = -z$ gives the average hedging loss.
2. $g_\sigma(z) = |z - \mathbb{E}[z]|$ gives the absolute average hedging error.
3. $g_\eta(z) = -z 1_{z < q_z(\beta)}$ gives the expected shortfall of the hedging loss, where the quantiles q are given by $\beta = \mathbb{P}[z \leq q_z(\beta)]$ for some $\beta \in (0, 1)$.

Remark 4.3.2. In Example 4.3.1 the first function g_μ corresponds to a loss function $L_t^{g_\mu}$ which is not necessarily non-negative. This violates the definition of loss functions given by Definition 2.1.3. However, this is not a huge problem, since we could equivalently consider the value of θ which minimises

$$\int (\mathbb{E}[-E_T(\theta, \theta') | \mathcal{F}_t] + K) p(\theta | V) d\theta$$

for some large enough $K > 0$ (specifically $K \geq -\inf_{\theta \in \Theta} \mathbb{E}[E_T(\theta, \theta') | \mathcal{F}_t]$). The adjusted loss function

$$\tilde{L}^g(\theta, \theta') = \mathbb{E}^\theta[g(E_T(\theta, \theta')) | \mathcal{F}_0] + K$$

is now non-negative and the minimising parameter is also the solution to the original minimisation problem. However, now we can no longer guarantee that $\tilde{L}(\theta, \theta) = 0$ so we violate the definition again. Hence we continue to use L^g rather than L but remark that the positivity violation is a technical and not conceptual problem. For the case of hedging, deviations from the true parameter can be both bad and good — profitable and non-profitable, rather than bad in either case. Hence $L(\theta, \theta') < 0$ simply means that there are alternative parameter values θ' which are more profitable than parameter θ . The violation of the definition is necessary in this financial context because, typically, we wish to capture as much (hedging) profit as possible.

To demonstrate our optimal Bayesian hedging methodology we concentrate in this chapter on the delta hedging strategy discussed in the previous section. As mentioned, this requires holding a portfolio with value Π of Δ units of stock S and Ψ units of cash B , so that a small movement in the underlying S causes an equal change in Π as it does in the value f_t^X of the option X we are hedging. We choose the delta hedge because it is the simplest, most widely understood, and most commonly used hedge. Furthermore, the delta hedge is vulnerable to large movements in the underlying S and volatility of the underlying, so mis-specification of the volatility can lead to large hedging errors. Since we use the local volatility model throughout this thesis as our base example, the delta hedging strategy is ideal for examining whether our Bayesian hedges really can reduce the vulnerability of the strategy to model mis-estimation and mis-specification.

For the delta strategy defined by (4.5), the loss functions in Example 4.3.1 give three very different Bayesian hedges. We will call the first the μ -hedge and denote the corresponding Bayesian hedge parameter by θ_t^μ , which is shorthand for $\theta_{L^{g_\mu}, p_t}(V)$ using the notation of (4.7), where $p_t(\theta|V) = p(\theta|(V_s)_{0 \leq s \leq t})$ is the probability density of θ corresponding to probability measure Q_t . Q_t is the Bayesian posterior probability measure given all market price observations up to time t . The μ -hedge minimises the expected relative hedging loss, or equivalently, maximises the expected relative hedging profit. The expectation is a double expectation

$$\theta_t^\mu = \theta_{L^{g_\mu}, p_t}(V) = \arg \min_{\theta'} \left\{ E^{Q_t} [\mathbb{E}^\theta [-E_T(\theta, \theta') | \mathcal{F}_t]] \right\},$$

taken over different paths of S in the inner expectation and different values of θ in the outer expectation using the Bayesian posterior density Q_t at time t given by (4.6).

The second hedge, θ_t^σ , we will refer to as the σ -hedge. It minimises the expected absolute relative hedging error, where again the inner expectation is taken over the different paths of S and the outer expectation is taken with respect to the different values of θ using the Bayesian posterior Q_t :

$$\theta_t^\sigma = \theta_{L^{g_\sigma}, p_t}(V) = \arg \min_{\theta'} \left\{ E^{Q_t} [\mathbb{E}^\theta [|E_T(\theta, \theta') - \mathbb{E}^\theta [E_T(\theta, \theta')] | | \mathcal{F}_t]] \right\}$$

The third hedge we consider, θ_t^η , will be called the η -hedge. It is the hedge that minimises the expected value of the expected relative hedging loss in the worst β scenarios i.e. path evolutions. So

first the expected shortfall is computed over different paths of S and then the expectation of this is taken with respect to Q_t over different values of θ :

$$\theta_t^\eta = \theta_{L^{\theta_t}, p_t}(V) = \arg \min_{\theta'} \left\{ E^{Q_t} [\mathbb{E}^\theta [-E_T(\theta, \theta') 1_{E_T(\theta, \theta') < q(\beta)} | \mathcal{F}_t]] \right\}$$

This hedge is one a more risk-averse agent might favour over the μ -hedge.

4.3.1 Hedging Improvement

As agents selling options and hedging, we are interested in improving the hedging performance of our hedging portfolios. The advantage of the Bayesian approach is that the expected improvement (i.e. reduction) of the loss L (with respect to the Bayesian posterior) is positive. To see this, let θ^0 be the original (e.g. MAP) parameter used for hedging. Then the improvement in the hedging performance corresponding to loss function L using the Bayesian strategy θ_L is

$$I(\theta^0, \theta_L) := L(\theta^*, \theta^0) - L(\theta^*, \theta_L), \quad (4.10)$$

where θ^* is the true parameter value. In practice we do not know θ^* ; we compute the expected value (with respect to the posterior density $p(\theta|V)$) of the improvement $I(\theta^0, \theta_L)$:

$$\begin{aligned} \mathbb{E}[I(\theta^0, \theta_L)] &= \int L(\theta, \theta^0) p(\theta|V) d\theta - \int L(\theta, \theta_L) p(\theta|V) d\theta \\ &\geq 0, \end{aligned} \quad (4.11)$$

by the definition of the Bayes estimator in (4.7).

Furthermore, we can calculate the variance of (4.10),

$$\mathbb{V}[I(\theta^0, \theta_L)] = \int [L(\theta, \theta^0) - L(\theta, \theta_L)]^2 p(\theta|V) d\theta - \{\mathbb{E}[I(\theta^0, \theta_L)]\}^2, \quad (4.12)$$

to give confidence intervals for the actual improvement (4.10). For example,

$$[\mathbb{E}[I(\theta^0, \theta_L)] - 2\sqrt{\mathbb{V}[I(\theta^0, \theta_L)]}, \mathbb{E}[I(\theta^0, \theta_L)] + 2\sqrt{\mathbb{V}[I(\theta^0, \theta_L)]}] \quad (4.13)$$

would correspond to a 95% confidence interval around the mean (4.11) if we approximate the distribution of (4.10) as Gaussian. If the variance (4.12) is low, then we can get fairly tight bounds on the actual difference.

$\mathbb{E}[I(\theta^0, \theta_L)] \geq 0$ might seem a trivial (or tautologous) result, but the implications are fundamental to the motivation behind the Bayesian approach. Indeed, because we can actually calculate the expected difference (4.11), if it is found to be large, then there is a good chance the actual hedging improvement (4.10) is significant. Of course, how close the two quantities (4.11) and (4.10) are to one another will depend on the accuracy of the posterior density function $p(\theta|V)$. To this end, we have shown in Chapter 2 that, under particular assumptions on the parameter space Θ and pricing functions f , if a true model parameter θ^* exists, then

$$p(\theta|V) \rightarrow \delta(\theta - \theta^*)$$

in probability as the number of observations V increases (where $\delta(z)$ is corresponds to the Dirac measure — zero everywhere except at $z = 0$).

Note however that in our simulations we adjust the posterior according to the newly observed calibration option prices each day and then change our MAP and Bayesian hedges accordingly. Hence, the expectations and confidence intervals we compute do not estimate the quantities we are actually simulating. Instead, they provide a rough quasi-approximation of the actual hedging improvements.

4.3.2 Derived Model Uncertainty Measures

For an agent selling a contract with value X and hedging to protect his position, the mean hedging error resulting from market risk of hedging the contract with value X using fixed model θ is a good prediction of the hedging loss to the agent that is attributable to selecting the wrong model for hedging and pricing. A function of the relative hedging error multiplied by the price of the option is thus a sensible and correctly-scaled measure of the model incorrectness. By sensible we mean that the difference between the true value of the contract and the corresponding replicating (hedging) portfolio under model θ accurately equates to the cost of using θ when it is not the correct model. And by correctly-scaled we mean that the monetary value of the hedging error equates with the monetary cost of using incorrect model θ to price the contract with value X .

With reference to Chapter 3, where we discussed model uncertainty measures, we can use hedging errors to construct further convex model uncertainty measures. We show in the following theorem that hedging error loss functions can be transformed to construct the convex penalty functionals used in the representation given by Theorem 3.3.6 for convex model uncertainty measures.

Theorem 4.3.3. *Take Θ finite, let $L^g(\theta, \theta')$ be the loss function (for a claim with value X) given by equations (4.8-4.9), and assume that the value $\Pi(S(\theta), \Delta(\theta'))$ of the hedging strategy is concave in θ' and g is decreasing and convex. Then the function $\mu^*(X)$ defined in (3.17) by*

$$\mu^*(X) = \inf_x \sup_{Q \in \mathcal{Q}_{\mu^*}} \{E^Q[2|\mathbb{E}^\theta[X] - x|] - \alpha(Q)\},$$

where $\mathcal{Q}_{\mu^*} = \overline{\{\delta_\theta : \theta \in \Theta\}}$ is the convex closure of the set of atomic measures for models in Θ and penalty functional α is given by

$$\alpha(Q) = \alpha \left(\sum_{i=1}^{|\Theta|} \lambda_i \delta_{\theta_i} \right) = \alpha \left(\delta_{\sum_{i=1}^{|\Theta|} \lambda_i \theta_i} \right)$$

for some weights $\lambda_1, \dots, \lambda_{|\Theta|}$ summing to 1 where $Q = \sum_{i=1}^{|\Theta|} \lambda_i \theta_i$ and

$$\alpha(\delta_{\theta'}) = \int_{\theta} L^g(\theta, \theta') f_0^X(\theta) p(\theta|V) d\theta, \quad (4.14)$$

is a convex model uncertainty measure, i.e. satisfies Definition 3.3.4.

Proof. We need to show that α is convex:

$$\begin{aligned}
& \alpha(\lambda\delta_{\theta'_1} + (1-\lambda)\delta_{\theta'_2}) \\
&= \int_{\theta} \mathbb{E}[g([\Pi(S(\theta), \Delta(\lambda\theta'_1 + (1-\lambda)\theta'_2)) - h(S_T)]/f_0^X(\theta))] f_0^X(\theta)p(\theta|V) d\theta \\
&\leq \int_{\theta} \mathbb{E}[g([\lambda\Pi(S(\theta), \Delta(\theta'_1)) + (1-\lambda)\Pi(S(\theta), \Delta(\theta'_2)) - h(S_T)]/f_0^X(\theta))] f_0^X(\theta)p(\theta|V) d\theta \\
&\leq \int_{\theta} \mathbb{E}[\lambda g([\Pi(S(\theta), \Delta(\theta'_1)) - h(S_T)]/f_0^X(\theta)) + (1-\lambda)g([\Pi(S(\theta), \Delta(\theta'_2)) - h(S_T)]/f_0^X(\theta))] \\
&\quad \times f_0^X(\theta)p(\theta|V) d\theta \\
&= \lambda\alpha(\delta_{\theta'_1}) + (1-\lambda)\alpha(\delta_{\theta'_2}),
\end{aligned}$$

which is the definition of convexity of α . The first inequality occurs because Π is concave in θ' and g is decreasing. The second inequality results from g being convex. \square

Example 4.3.4. $g_\mu(z) = -z$ defined in Example 4.3.1 is both decreasing and convex, and hence, if $\Pi(S(\theta), \Delta(\theta'))$ is concave in θ' , gives a convex model uncertainty measure by Theorem 4.3.3 above. However, g_σ is not decreasing and g_η is not convex, so neither necessarily gives a convex model uncertainty measure.

Observe by Theorem 4.3.3 that the skew of Π plays a role in determining the convexity of the penalty function. So it is worthwhile spending some time investigating whether Π is concave or convex in θ' (if either). Although in general closed-form formulas do not exist for the values of hedging profits, we do have one for the case of delta hedging when the asset price S follows geometric Brownian motion.

Suppose the correct model for S is given by

$$dS_t = a_t S_t dt + \sigma_t S_t dZ_t$$

but we hedge using the Black-Scholes model

$$dS_t = a' S_t dt + \sigma' S_t dZ_t,$$

where a_t, σ_t are \mathcal{F}_t -adapted processes and a', σ' are constants. Note that σ_t is the realised instantaneous volatility — so we could be in a Black-Scholes model (in which case σ_t is constant) or a local volatility model, or a stochastic model. Then El Karoui et al. [54], and later Davis [42], have shown that the relative terminal hedging error from a continuous delta hedging strategy, with value Π_t for shorting an option X , with payoff $h(S_T)$ at time T , initial value f_0^X and price function f_t^X at time t , is given by

$$E_T(\sigma, \sigma') = \frac{\Pi_T - h(S_T)}{f_0^X} = \frac{1}{2f_0^X} \int_0^T e^{r(T-t)} S_t^2 \Gamma_t ((\sigma')^2 - \sigma_t^2) dt, \quad (4.15)$$

where the *gamma* $\Gamma_t = \partial^2 f_t^X / \partial S_t^2$ is the second derivative of the price with respect to the underlying. Hence, twice differentiating equation (4.15) with respect to σ' gives

$$\frac{\partial^2 \Pi_T}{(\partial \sigma')^2} = \int_0^T e^{r(T-t)} S_t^2 \Gamma_t dt, \quad (4.16)$$

since S_t and Γ_t are functions of the true volatility σ_t not σ' .

Now, since $e^{r(T-t)}$ and S_t are always non-negative, the skew of Π_T is determined purely by the gamma Γ_t . (4.16) tells us that if $\Gamma_t \leq 0$, then Π_T is concave in σ' , but if $\Gamma_t \geq 0$, then Π_T is convex in σ' . Of course if Γ_t switches signs through the lifetime of the option, then the skew of Π_T will depend on the final sign of the integral in (4.16).

For example, $\Gamma_t \geq 0$ always for standard European calls and puts. So if we go long a European call instead of shorting it, then the gamma becomes non-positive and Π_T becomes concave, so we satisfy the assumption of Theorem 4.3.3. Hence, using the loss function L_{g_μ} , for example, with this hedge will give a convex penalty functional because, as noted in Example 4.3.4, g_μ is decreasing and convex.

On the other hand, if we short a European call, then the gamma is non-negative, so Π_T becomes convex. One might suggest that we could still construct penalty functionals by choosing g increasing and convex instead. However, if g were increasing, then we would be assigning greater loss to higher hedge profits than lower ones, which is nonsensical.

4.3.3 Relationship with Utility Functions

A *utility function* U is a map from $\mathbb{R} \rightarrow [-\infty, \infty)$ representing an agent's preferences over different contingent claims. More precisely, we suppose that the agent's preferences are expressed through an expected utility representation of the form

$$Y \text{ is preferred to } X \iff \mathbb{E}[U(Y)] \geq \mathbb{E}[U(X)]$$

where X and Y are contingent claims and U is increasing and concave. A large proportion of mathematical finance literature has been devoted to using the theory of utility functions in asset pricing and portfolio optimisation problems (see for example [81]). U is increasing because (typically) more wealth is preferred to less and concave because (usually) an agent obtains less additional utility from each additional unit of wealth. See Föllmer & Schied's book [67] for a full introduction.

In the context of optimal hedging strategies, the utility theory approach would be to maximise the expected utility of the relative hedging profit, i.e.

$$\text{Find } \theta' \text{ which maximises } \mathbb{E}^{Q,\theta} [U([\Pi(\theta, \theta', \omega) - h(S(\theta, \omega))] / f_0(\theta))], \quad (4.17)$$

where recall Π is the value of the hedging strategy, h is the payoff and $f_0(\theta)$ is the initial value of the option we are hedging. Observe the expectation is taken over the different paths ω (using measure

θ) of the driving process and also over the different possible models θ (using measure Q). On the other hand, in the context of this chapter, the Bayesian approach is to

$$\begin{aligned} \text{Find } \theta' \text{ which minimises} & \int L(\theta, \theta') p(\theta|V) d\theta \\ = & \int \mathbb{E}^\theta [g([\Pi(\theta, \theta', \omega) - h(S(\theta, \omega))]/f_0(\theta))] p(\theta|V) d\theta \\ = & \mathbb{E}^{Q,\theta} [g([\Pi(\theta, \theta', \omega) - h(S(\theta, \omega))]/f_0(\theta))], \end{aligned}$$

$$\text{i.e. Find } \theta' \text{ which maximises } \mathbb{E}^{Q,\theta} [-g([\Pi(\theta, \theta', \omega) - h(S(\theta, \omega))]/f_0(\theta))]. \quad (4.18)$$

So we see that the utility approach (4.17) and Bayesian approach (4.18) coincide precisely when

$$U = -g. \quad (4.19)$$

When this equality holds, we must have then that g is decreasing and convex (since U is increasing and concave). Berger finds the same identity in his book [11]. Föllmer & Schied [67] arrive at a very similar identity, although they take supremums of U and g with respect to the unknown model/measure θ rather than expectations.

The identity is a satisfying theoretical result because decreasing and convex were precisely the properties required of g in Theorem 4.3.3 for the corresponding measure μ_β^g to be a convex model uncertainty measure. This appears to add some completeness to the Bayesian theory and provide further cohesion between the traditional expected utility approach in mathematical finance and the proposed Bayesian approach of this thesis.

4.4 Numerical Examples

We continue the application of our methods to the implementation of the local volatility model. We consider two examples. In the first, the daily observed underlying asset and calibration prices are produced by the local volatility surface given in [89], which we have worked with extensively in the previous two chapters. In the second example, the daily observed underlying asset and calibration prices are generated using a Heston stochastic volatility model, in a similar fashion to Hull & Suo [86]. The reasons for choosing the Heston model instead of another model or even real market data is discussed in the next section.

For each example we calibrate local volatility models using the formulation and sampling methods of Chapter 2. From the resulting posterior distribution of calibrated surfaces we compute the three Bayesian delta-neutral hedges (μ -, σ - and η -) and the ‘true’ and MAP delta hedges and MAP delta-vega hedge for comparison. By true hedge we mean the delta hedge found using the true local volatility surface that we are simulating the paths on. We choose the true local volatility delta (local delta) hedge rather than the Black-Scholes implied delta hedge following the findings of Crepey [39].

In his paper Crepey concludes that the local delta of an option should be preferred to the Black-Scholes implied delta for hedging. He demonstrates this using numerical tests and shows his claim to be true even when transaction costs are included.

For both examples we record the terminal relative hedging error distributions for options after daily re-hedging and re-calibration to the new asset and calibration prices (calculated on the true model we are trying to recover). Daily re-calibration is what is done in practice by traders [86], so it is the best way of testing the effectiveness of our Bayesian hedges.

We remark here that we have deliberately focused on comparing the Bayesian hedges to hedging on the MAP surface and excluded comparison with other surfaces. In particular, the local volatility surface constructed using Dupire's formula (2.15) is omitted from our investigations. The Dupire formula is an explicit method for reconstructing the local volatility surface given the values of European call prices for all strikes and maturities. But as we have discussed in Section 2.3.2, a full complement of prices does not exist, and application of Dupire's formula to a handful of noisy prices gives non-smooth solutions. Hence some form of regularisation and interpolation is always necessary with the Dupire formula. The cost of calculating derivatives for the formula, however, makes this process expensive, and it becomes much more efficient to use the regularised calibration-error minimisation routines — of [98], [89] and [28] for example — which, as we have mentioned in Remark 2.4.1, is equivalent to maximising our Bayesian posterior density to get the MAP estimate.

We again acknowledge that the MAP surface we have so far taken for granted as being available to us, by selecting the sample of the posterior which maximises the posterior, may not be the true MAP surface. We are repeating the point made in Chapter 2 by Remark 2.5.5, but it is important enough to warrant repetition. The true MAP surface (found by maximising the posterior) is typically not computationally easy/quick to find, and hence we settle for selecting the sample which has maximum posterior. This may be argued to make comparison less true and less *fair*. But recall that the calibration error of our MAP surface is comparable to that found in minimising regularised error functionals (Remark 2.5.5) by other papers, e.g. [89]. Furthermore, we must remember that each Bayesian strategy is itself not the true Bayesian hedge, but only an approximation based on the available samples. But we showed in Section 2.5.5 that the Metropolis sampling algorithm converged to within the acceptable PSRF values.

4.4.1 Heston Model

The Heston model [82] is a mean-reverting stochastic volatility model of the type studied by Cox et al. in interest rate models [36]. It is given by the equations

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma_t S_t dW_t, \\ d[\sigma_t^2] &= \alpha(\beta - \sigma_t^2) + \gamma \sigma_t dZ_t, \end{aligned} \tag{4.20}$$

where $(S_t)_{t \geq 0}$ is the value process of the underlying asset S with time-0 value S_0 and $(\sigma_t^2)_{t \geq 0}$ is the value process of the variance of S and has time-0 value σ_0^2 . The drivers of the processes, $(W_t)_{t \geq 0}$ and $(Z_t)_{t \geq 0}$, are Brownian motions with respect to the physical measure \mathbb{P} and have correlation coefficient ρ . In this model, μ is the drift of the asset price S , α is the rate of mean reversion of the volatility process σ^2 , β is the long-run mean of the volatility process σ^2 , and γ is the volatility of the volatility.

Observe that when $\gamma = 0$, the volatility is a deterministic process of time t :

$$\sigma_t = [\beta - (\beta - \sigma_0^2)e^{-\alpha t}]^{1/2},$$

and hence reduces to the local volatility model $\sigma(S, t)$ given by (2.14) in Chapter 2 (albeit without dependence on the asset price S).

However, for the purposes of our investigation we want the distribution of paths and structure of prices generated by the Heston model to be dissimilar from those produced by a local volatility model. This will test the robustness of our method to mis-specification of the model. For the purposes of the numerical test, we borrow the setup of the S&P 500 example of [28] that we studied in Chapter 2 and Chapter 3. In particular we take

$$\begin{aligned} S_0 &= \$590 & \sigma_0^2 &= 0.012 & r &= 0.06 & d &= 0.0262 & \mu &= r - d \\ \alpha &= 5.3 & \beta &= 0.018 & \gamma &= 0.25 & \rho &= -1, \end{aligned}$$

and use the same set of strikes and maturities given in the second table in Appendix A.1. The values of α , β , γ , and ρ were estimated from the dataset given in Appendix A.1 using a basic non-linear least squares optimiser — it is not important that we closely reproduce the prices given by [28].

Generating the paths following the Heston process given by (4.20) is simple — two discretised ρ -correlated Brownian paths are simulated via the Euler scheme and the values for S_t and σ_t^2 are updated by increments of time. Positivity of σ_t^2 is ensured by using

$$\sigma_{t+\delta t}^2 = |\sigma_t^2 + \alpha(\beta - \sigma_t^2) + \gamma\sigma_t dZ_t|.$$

In order to calibrate local volatility models, we need to generate the corresponding market prices, and this is less straightforward. In his original paper [82], Heston found a closed form solution using PDE methods for the value of a European call option which involves the computation of an integral. Subsequent improvements have been made to Heston's formula, and in this thesis we use the Fast Fourier Transform method introduced by Carr & Madan [23]. Full details of the pricing formula are given in Appendix B.2.

Finally, in this section we justify our choice of the Heston Model. We could have used a different stochastic volatility model — Hull-White [88] or Log Ornstein-Uhlenbeck [127] for example. We could have used a different class of model entirely — jump-diffusion model or general Lévy process. Or alternatively we could have experimented with real market data. However, there is a strong reason for rejecting each of these alternatives.

Firstly, it would not have been viable to use real market data because of the relative dearth of available data. For a particular asset over a particular time interval, we have only one asset price evolution and one evolution of option prices to calibrate against. However, to calculate hedging profit distributions we need thousands of evolutions. We could take non-overlapping intervals of the same asset price but, as Elder points out in his thesis [55], there are not enough intervals. For example, 1000 runs of hedging a 3 month claim necessitates 250 years worth of data, but the FTSE 100 is only 25 years old and the NASDAQ is only 38 years old.

Secondly, much of the empirical work carried out on asset returns and implied volatilities reveals that stochastic volatility is a good model for capturing the observed properties. Cont [29] observes heavy tails in asset returns, asymmetry of large downward movements and smaller upward movements, volatility clustering, strong negative correlation between volatility and asset price — all of which can be demonstrated by a stochastic volatility model. Rebonato [116] and Shephard [125] recommend stochastic volatility models for their ability to capture the skews and kurtosis seen in the implied volatility surfaces of foreign exchange and equity prices. However, the papers concede that stochastic volatility models, unlike jump processes, cannot capture short-term skews, i.e. do not give fat tails for options with short maturities. But Cont & Tankov [32] further point out that, unlike stochastic volatility models, time-homogeneous jump models and independent-increments exponential-Lévy processes cannot reproduce observed implied volatility term structures.

Thirdly, the choice of Heston model over other stochastic volatility models is one of convenience and mathematical tractability. Heston [82] finds a closed-form solution for European call prices for all values of correlations coefficient ρ — in contrast to the Hull-White model that can only be analytically solved for $\rho = 0$, and the Log Ornstein-Uhlenbeck model which has no known closed-form solution for European call prices for any value of ρ .

4.4.2 Dynamic Hedging Strategies

Following the setup in the introduction, we consider two (dynamic) hedging strategies: a) a portfolio of the underlying S and cash B ; b) a portfolio of the underlying S and cash B and option Y . For the first case we shall form a delta-neutral portfolio and in the second case we will construct a delta-vega-neutral portfolio.

Consider first the delta-neutral portfolio. At time t the hedging portfolio has value Π_t and is composed of Δ_t of stock S and $\Pi_t - \Delta_t S_t$ of cash B . Π is used to hedge an option X written on S with payoff $h^X(S)$ at maturity time T and has initial value $\Pi_0 = f_0^X$ at initial time 0. Then at time $t \in [0, T]$ the relative hedging error E_t is given by

$$\begin{aligned} f_0^X E_t &= \Pi_t - f_t^X \\ &= \Pi_0 + \int_0^t \Delta_u dS_u + \int_0^t q\Delta_u S_u du + \int_0^t (\Pi_u - \Delta_u S_u) dB_u - f_t^X, \end{aligned} \quad (4.21)$$

where q is the annualised continuous dividend yield.

In particular, when the stock evolves according to one model, i.e. $S = S(\theta)$, and the hedging strategy is calculated using a different model, e.g. $\Delta = \Delta(\theta')$, then the relative hedging error $E_t(\theta, \theta')$ at time t on any single path realisation is given by

$$\begin{aligned} f_0^X(\theta) E_t(\theta, \theta') &= \Pi_t(S(\theta), \Delta(\theta')) - f_t^X(\theta) \\ &= \Pi_0(\theta') + \int_0^t \Delta_u(\theta') dS_u(\theta) + \int_0^t q\Delta_u(\theta') S_u(\theta) du \\ &\quad + \int_0^t (\Pi_u - \Delta_u(\theta') S_u(\theta)) dB_u - f_t^X(\theta), \end{aligned}$$

where $f_t^X(\theta)$ is the price of option X at time t computed using model θ . For the portfolio to be delta-neutral, we must take $\Delta_t = \frac{\partial f_t^X}{\partial S_t}$ (see for example [16] for details).

In practice we cannot re-hedge continuously (and would not want to because of non-zero transaction costs for the purchases/sales of the underlying S), but instead re-hedge at discrete times $0 = t_0 < t_1 < \dots < t_{n-1}$ (and label $t_n = T$). Assuming now that the cash process B evolves as $dB_t = rdt$, the terminal relative hedging error $E_T(\theta, \theta')$ at maturity time T of the discrete delta hedging strategy is given by

$$\begin{aligned} f_0^X(\theta) E_T(\theta, \theta') &= \Pi_T(S(\theta), \Delta(\theta')) - h^X(S(\theta)) \\ &= f_0^X(\theta') e^{rT} + \sum_{i=0}^n [\Delta_{t_{i-1}}(\theta') (1 + q(t_i - t_{i-1})) - \Delta_{t_i}(\theta')] S_{t_i}(\theta) e^{r(t_n - t_i)} \\ &\quad - h^X(S(\theta)), \end{aligned} \tag{4.22}$$

where $\Delta_{t_i}(\theta') = \frac{\partial f_{t_i}^X(\theta')}{\partial S_{t_i}(\theta')}$ for $i = 1, \dots, n-1$ and $\Delta_{t_{-1}}(\theta') = \Delta_{t_n}(\theta') = 0$ for all θ' . Full details of this result can be found in Appendix C.2.

Consider next the delta-vega-neutral portfolio. At time t the hedging portfolio has value Π_t and is composed of Δ_t of stock S , Ξ_t of option Y and $\Pi_t - \Delta_t S_t - \Xi_t V_t^Y$ of cash B . Again the portfolio is used to hedge an option X written on S with payoff $h^X(S)$ at maturity time T and market price f_t^X at time t . As before, the hedging portfolio has initial value $\Pi_0 = f_0^X$ at time 0. Similarly, option Y , written on S with payoff $h^Y(S)$ at maturity time T , has valuation $f_t^Y(\theta)$ in model θ at time t . Then at time $t \in [0, T]$ the relative hedging error E_t is given by

$$\begin{aligned} f_0^X E_t &= \Pi_t - f_t^X \\ &= \Pi_0 + \int_0^t \Delta_u dS_u + \int_0^t \Xi_u df_u^Y + \int_0^t q\Delta_u S_u du \\ &\quad + \int_0^t (\Pi_u - \Delta_u S_u - \Xi_u f_u^Y) dB_u - f_t^X. \end{aligned} \tag{4.23}$$

For the portfolio to be delta-neutral and vega-neutral we set the coefficients Δ and Ξ to be

$$\Xi_t = \frac{\partial f_t^X / \partial \sigma_t}{\partial f_t^Y / \partial \sigma_t}, \tag{4.24}$$

$$\Delta_t = \frac{\partial f_t^X}{\partial S_t} - \Xi_t \frac{\partial f_t^Y}{\partial S_t}, \tag{4.25}$$

where σ_t is the implied volatility of underlying S at time t . For a full derivation of this result see for example [116].

As for the case of delta-neutral hedging portfolios, we do not in practice re-hedge continuously but instead re-hedge at discrete times $0 = t_0 < t_1 < \dots < t_{n-1}$ (and label $t_n = T$). The terminal relative hedging error $E_T(\theta, \theta')$ at maturity T of the discrete approximate delta-vega-neutral hedging strategy is then given by

$$\begin{aligned} f_0^X(\theta) E_T(\theta, \theta') &= \Pi_T(S(\theta), \Delta(\theta')) - h^X(S(\theta)) \\ &= f_0^X(\theta') e^{rT} + \sum_{i=0}^n [\Delta_{t_{i-1}}(\theta')(1 + q(t_i - t_{i-1})) - \Delta_{t_i}(\theta')] S_{t_i}(\theta) e^{r(t_n - t_i)} \\ &\quad + \sum_{i=0}^n [\Xi_{t_{i-1}}(\theta') - \Xi_{t_i}(\theta')] f_{t_i}^Y(\theta) e^{r(t_n - t_i)} - h^X(S(\theta)), \end{aligned} \quad (4.26)$$

where $\Delta_{t_i}(\theta')$ and $\Xi_{t_i}(\theta')$ are given by (4.25) and (4.24) respectively for $i = 1, \dots, n-1$ and $\Xi_{t_{-1}}(\theta') = \Xi_{t_n}(\theta') = \Delta_{t_{-1}}(\theta') = \Delta_{t_n}(\theta') = 0$ for all θ' . Full details of this result can be found in Appendix C.3.

4.4.3 Computation of the Loss Functions

Given the distribution of calibrated parameters found in Chapter 2 we need to evaluate the loss functions corresponding to Example 4.3.1. In the discrete re-hedging case we compute these functions for different pairs (θ, θ') by Monte Carlo using the formulae

$$\begin{aligned} L^{g_\mu}(\theta, \theta') &= \frac{1}{M} \sum_{m=0}^M -E_T^{(m)}(\theta, \theta'), \\ L^{g_\sigma}(\theta, \theta') &= \frac{1}{M} \sum_{m=0}^M \left| [E_T^{(m)}(\theta, \theta')] - \frac{1}{M} \sum_{m=0}^M E_T^{(m)}(\theta, \theta') \right|, \\ L^{g_\eta}(\theta, \theta') &= \frac{1}{(1-\beta)M} \sum_{m=0}^M -E_T^{(m)}(\theta, \theta') 1_{z < q_z(\beta)}, \end{aligned}$$

where the random variable $z = E_T^{(m)}(\theta, \theta')$ and, for each path $S^{(m)}$, $E_T^{(m)}(\theta, \theta')$ is as given by (4.22) and (4.26) for the delta and delta-vega hedges respectively.

4.4.4 Bayesian Updating (Recalibration)

Recall that the Bayes estimator θ_t^μ is given by

$$\theta_t^\mu = \theta_{L^{g_\mu}, p_t}(V) = \arg \min_{\theta'} \left\{ \int L^{g_\mu}(\theta, \theta') p_t(\theta|V) d\theta \right\}.$$

We have shown in the previous section how to calculate the loss functions L^g for different functions g of the hedging profit $E_T(\theta, \theta')$. So all that remains in order to compute the optimal Bayesian hedges, θ_t^μ , θ_t^σ and θ_t^η , is to sample the posterior densities $p_t(\theta|V)$ for the hedging times $t = t_0, \dots, t_{n-1}$ and minimise the expected loss. However, this is not a straightforward procedure.

We described in Section 2.5.3 how to sample the posterior distribution at time $t = t_0 = 0$, and in theory we could repeat this procedure at times $t = t_1, \dots, t_{n-1}$. Unfortunately this is too computationally expensive to be practically useful (or implementable). First of all, the multiple-chain Markov Chain Monte-Carlo algorithm used to generate samples θ_k is very time-expensive. Secondly, after generating a new set of samples at time $t > 0$ we would have to recompute the loss functions $L^g(\theta, \theta')$ for the new pairs (θ_k, θ_l) , which is again very time-expensive. And all this would have to be repeated at every hedge time t_i on every Monte-Carlo path $S^{(m)}$ — which is computationally far too expensive.

Instead, we keep the samples $\theta_1, \dots, \theta_N$ we found at time t_0 using the Metropolis algorithm described in Section 2.5.3, but adjust their ‘Bayesian weights’ in the Monte-Carlo calculation. Recall from expression (2.34) in Chapter 2 that after the time t_0 calibration and sampling, for any function f of θ , the Bayes estimator of the integral $\int_\theta L(\theta) p_{t_0}(\theta|V_0) d\theta$ is given by

$$\frac{1}{N} \sum_{i=1}^N L(\theta_i).$$

Now at time t_1 we have observed a new set of prices V_{t_1} and re-calibrated our framework by *updating* the posterior to $p_{t_1}(\theta|V)$:

$$\begin{aligned} p_{t_1}(\theta|V) &\propto p_{t_0}(\theta|V) \exp\{G_{t_1}(\theta)\} \\ &\propto \exp\{-\frac{1}{2\delta^2}[\lambda\|\theta - \theta_0\|^2 + \Sigma_{i=0}^1 G_{t_i}(\theta)]\}, \end{aligned}$$

where recall from Section 2.4.1 that λ represents our prior confidence, $\|\cdot\|$ is the regularisation norm in the Bayesian prior, and that the calibration error function $G_t(\theta)$ is given in Section 2.4.2 by

$$G_t(\theta) = \frac{10^8}{S_t^2} \sum_{i \in I} w_i |f_t^{(i)}(\theta) - V_t^{(i)}|^2.$$

We refer to this method of adjusting the posterior when more data is observed as *Bayesian updating* [70].

Having found the new posterior densities of the original surfaces $\theta_1, \dots, \theta_N$ (which we do easily and quickly by solving the Dupire PDE using a Crank-Nicolson finite difference as described in Section 2.5.3), we want to estimate the integral $\int_\theta L(\theta) p_{t_1}(\theta|V) d\theta$ using the same samples $\theta_1, \dots, \theta_N$, i.e. we wish to find the weights $y_1^{(1)}, \dots, y_1^{(N)}$ such that

$$\sum_{i=1}^N L(\theta_i) y_1^{(i)} \approx \int_\theta L(\theta) p_{t_1}(\theta|V) d\theta. \quad (4.27)$$

We want to re-use the same samples $\theta_1, \dots, \theta_N$, because resampling using the Metropolis algorithm would computationally take too long — given each sampling takes 5 minutes and we have to repeat this 60 times for each path for 1000 paths.

We find the weights $y_1^{(1)}, \dots, y_1^{(N)}$ using the method of *importance sampling* (see for example [70] for a detailed introduction) explained below. Observe we can write the integral on the right of (4.27) as

$$\begin{aligned} \frac{\int_{\theta} L(\theta) p_{t_1}(\theta|V) d\theta}{1} &= \frac{\int_{\theta} L(\theta) \frac{p_{t_1}(\theta|V)}{p_{t_0}(\theta|V)} p_{t_0}(\theta|V) d\theta}{\int_{\theta} \frac{p_{t_1}(\theta|V)}{p_{t_0}(\theta|V)} p_{t_0}(\theta|V) d\theta} \\ &\approx \frac{\frac{1}{N} \sum_{i=1}^N L(\theta_i) \frac{p_{t_1}(\theta_i|V)}{p_{t_0}(\theta_i|V)}}{\frac{1}{N} \sum_{i=1}^N \frac{p_{t_1}(\theta_i|V)}{p_{t_0}(\theta_i|V)}}. \end{aligned}$$

Hence we set

$$y_1^{(i)} = \left(\sum_{j=1}^N \frac{p_{t_1}(\theta_j|V)}{p_{t_0}(\theta_j|V)} \frac{1}{N} \right)^{-1} \frac{p_{t_1}(\theta_i|V)}{p_{t_0}(\theta_i|V)} \frac{1}{N},$$

and generally for any hedge time t_k , $k = 1, \dots, n-1$, we have the recursive formula for the weights

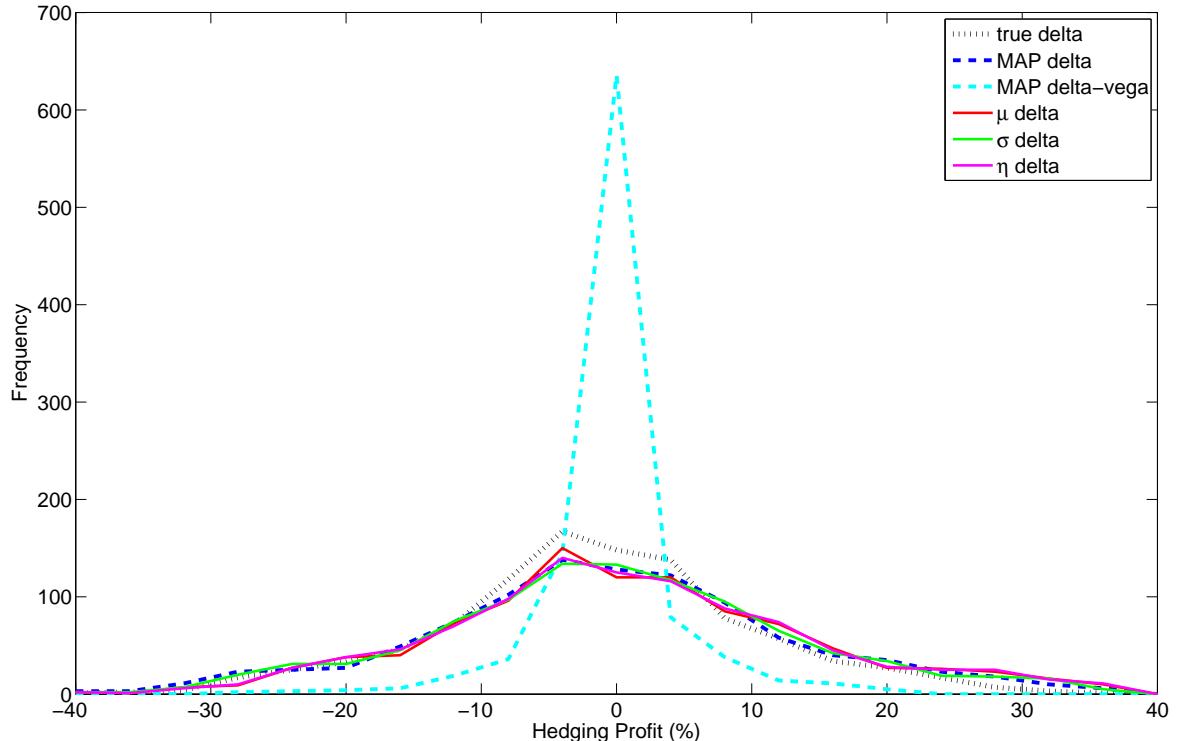
$$y_k^{(i)} = \left(\sum_{j=1}^N \frac{p_{t_k}(\theta_j|V)}{p_{t_{k-1}}(\theta_j|V)} y_{k-1}^{(j)} \right)^{-1} \frac{p_{t_k}(\theta_i|V)}{p_{t_{k-1}}(\theta_i|V)} y_{k-1}^{(i)}. \quad (4.28)$$

Note that the recursion (4.28) has to be carried out on every path $S^{(m)}(\theta^*)$ simulated on the true surface θ^* .

4.4.5 Hedging Results

We simulate paths in a local volatility model for the first test case, and then simulate paths in a Heston model for the second case. For each example we hedge two options: a 3-month at-the-money European call option and a 3-month up-and-out barrier call option. This allows us to assess the impact of our new Bayesian hedging strategies on a liquid vanilla option and an exotic option respectively [55]. Akin to Crepey's [39] numerical experiments, we run 1000 simulations for each test. Below each histogram of hedging profits we table the mean, absolute deviation, and expected shortfall of the hedging profits for the MAP strategies and the Bayesian strategies. We compute the improvement in each hedging strategy and also give the quasi 95% confidence intervals, calculated by (4.13), for these improvements. For the MAP delta-vega hedge we use a European call option with maturity 3 months and strike $0.99S_0$.

The results of the hedging distributions for both options are very encouraging. As we would expect for the European call option, the delta-vega hedge is far superior to any of the delta hedges. However, comparing the delta strategies we see that the Bayes hedges can capture some improvement, 1.3% in the case of mean hedge profit and 3.6% in the case of 5% profit shortfall. These are rather modest numbers. But also note that the improvement in the absolute deviation and shortfall lie within the quasi predicted intervals. The mean hedge profit falls just outside the quasi predicted confidence interval, however.



	mean hedge profit	absolute deviation	5% profit shortfall
true delta hedge	0.4	12.0	-26.2
MAP delta hedge	1.5	14.3	-29.6
MAP delta-vega hedge	1.4	5.1	-12.5
μ - delta hedge	2.8	14.3	-26.1
σ - delta hedge	1.9	14.3	-27.9
η - delta hedge	2.7	14.3	-26.0
actual hedging improvement	+1.3	-0.0	+3.6
quasi predicted conf. int.	[1.1,1.2]	[-0.1,0.3]	[0.6,4.1]

Figure 4.3: For local volatility underlying: (upper) histograms of hedging profits, when the true model is local volatility, for European call option with strike 5000 ($S_0 = 5000$) and maturity 3 months for the correct (true) delta hedging strategy and the MAP (delta and delta-vega) strategies and Bayes delta strategies. (lower) Table of key statistics for each hedging strategy and the realised improvements and 95% quasi confidence intervals for the improvements of these statistics made by the Bayes hedges over the MAP delta hedge.

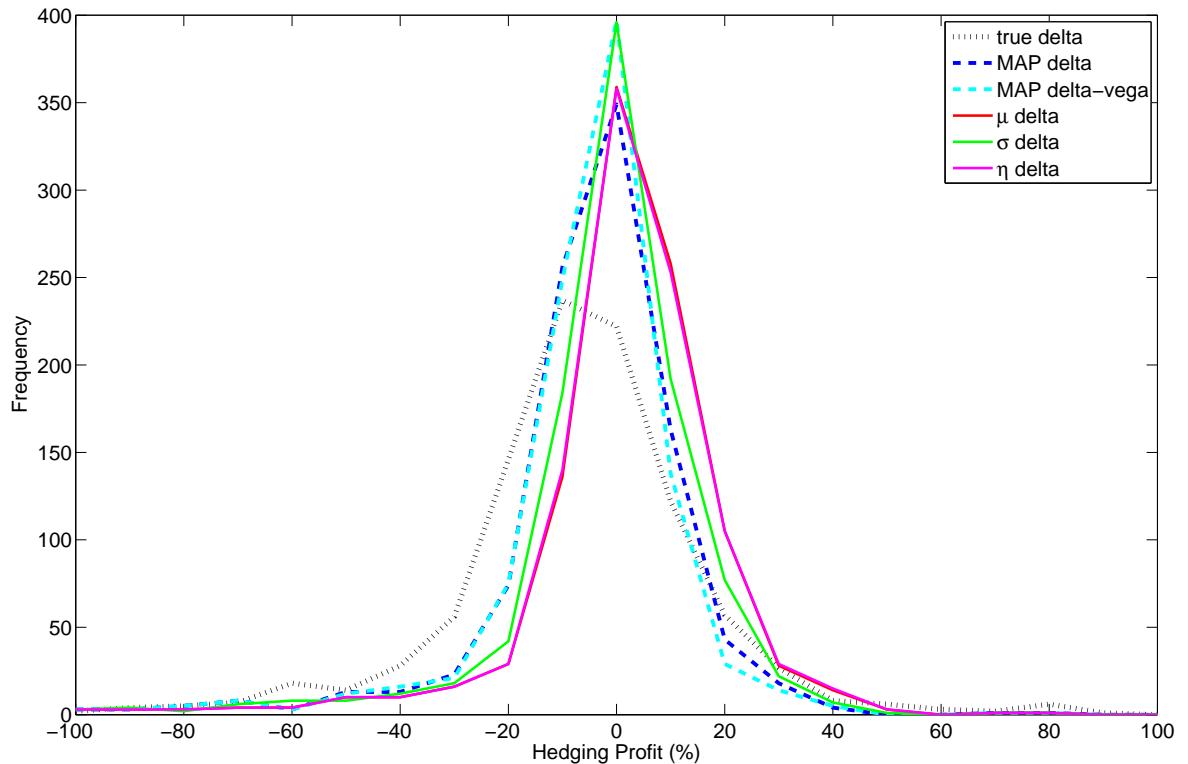


Figure 4.4: For local volatility underlying: (upper) histograms of hedging profits, when the true model is local volatility, for up-and-out barrier call option with strike 5000 ($S_0 = 5000$), barrier 5500 and maturity 3 months for the correct (true) delta hedging strategy and the MAP (delta and delta-vega) strategies and Bayes delta strategies. (lower) Table of key statistics for each hedging strategy and the realised improvements and 95% quasi confidence intervals for the improvements of these statistics made by the Bayes hedges over the MAP delta hedge.

The hedging statistics for the barrier option are far more interesting. We see that the hedging distributions for the μ -hedge and η -hedge are very similar and that the improvements are vast. The Bayes hedges increase the mean relative hedge profit and 5% profit shortfall by 7.2% and 22.1% respectively. These Bayes delta hedges even outperform the delta-vega hedge. The improvement in the absolute deviation of the hedging profits is negligible. Moreover, all improvements lie within the predicted confidence intervals. But the intervals for absolute deviation and profit shortfall are very wide — over 300% and 150% respectively. This is probably caused by the discontinuous payoff of the up-and-out barrier call option — very in-the-money just below the barrier and zero at the barrier. This causes large deltas at the barrier and so, for some paths, the delta hedges make large losses/profits. Hull & Suo [86] have similarly observed in their numerical tests that there is significant model uncertainty when the local volatility model is used to price barrier options and that this is an important result.

The hedging results for paths simulated in the Heston model are largely similar. For the European call option, Figure 4.5 shows that the MAP delta-vega hedge again vastly outperforms the delta hedges and the improvements using the Bayesian hedges are fairly modest, 3.2% and 4.8% for mean relative hedge profit and 5% shortfall respectively. The 0.4% reduction in absolute deviation is negligible. Again, the absolute deviation and expected shortfall improvements both lie within their reasonably tight quasi confidence intervals but the mean hedge profit lies just outside its quasi predicted confidence interval. For the barrier option, however, Figure 4.6 shows something very interesting. Again we see the Bayes hedges dramatically increase the mean hedge profit and 5% profit shortfall, 12.8% and 32.6% respectively, and hardly changes the absolute deviation. All the changes lie within their quasi predicted confidence intervals too, although the bounds for absolute deviation and 5% profit shortfall are again very wide. However, this time we see that the hedging profit distributions for the Bayesian μ -hedge and η -hedge almost replicate the hedging profit distribution of the MAP delta-vega hedge.

The graphs of this section have shown that for vanilla options the Bayesian hedges can only modestly improve the hedging performance of the delta hedge, and that the delta-vega hedge remains vastly superior. However, for exotic options, such as the barrier option, the Bayesian delta μ -hedge and η -hedge can improve average returns by over 15% — and this is for a short-dated option of 3 months — and also outperform the delta-vega hedges. Once we factor in transaction costs, which can be very high for vega hedging, e.g. Gondzio et al. [76] assume 2.5% transaction costs, we see that the improvements of the Bayes delta hedges are even more impressive for the barrier option and now potentially significant for the European call option.

Secondly, we have seen a good fit of actual hedging improvements to the quasi predicted intervals. However, this fit does not always hold and the intervals not always tight or helpful. This should not be surprising, however. Recall that formula (4.13) computes the expected improvement using the

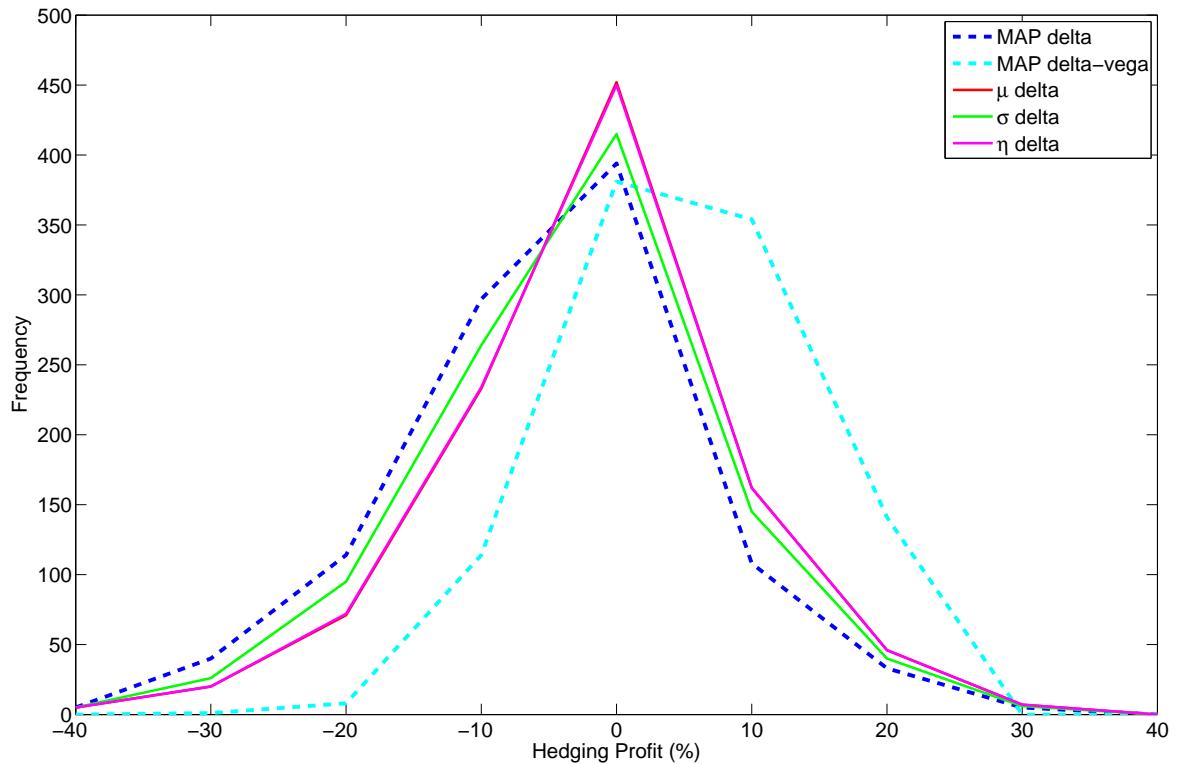


Figure 4.5: For Heston underlying: (upper) histograms of hedging profits, when the true model is Heston stochastic volatility, for European call option with strike \$590 ($S_0 = \590) and maturity 3 months for the MAP (delta and delta-vega) strategies and Bayes delta strategies. (lower) Table of key statistics for each hedging strategy and the realised improvements and 95% quasi confidence intervals for the improvements of these statistics made by the Bayes hedges over the MAP delta hedge.

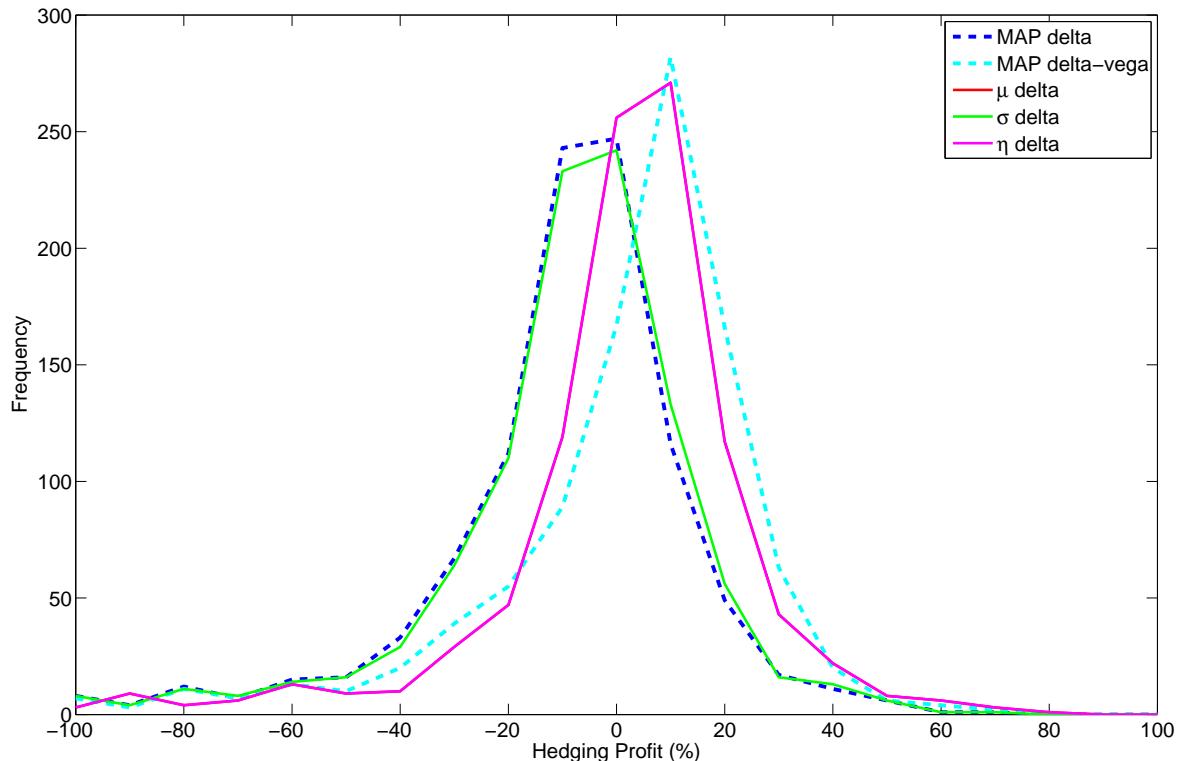


Figure 4.6: For Heston underlying: (upper) histograms of hedging profits, when the true model is Heston stochastic volatility, for up-and-out barrier call option with strike \$590 ($S_0 = \590), barrier \$650 and maturity 3 months for the MAP (delta and delta-vega) strategies and Bayes delta strategies. (lower) Table of key statistics for each hedging strategy and the realised improvements and 95% quasi confidence intervals for the improvements of these statistics made by the Bayes hedges over the MAP delta hedge.

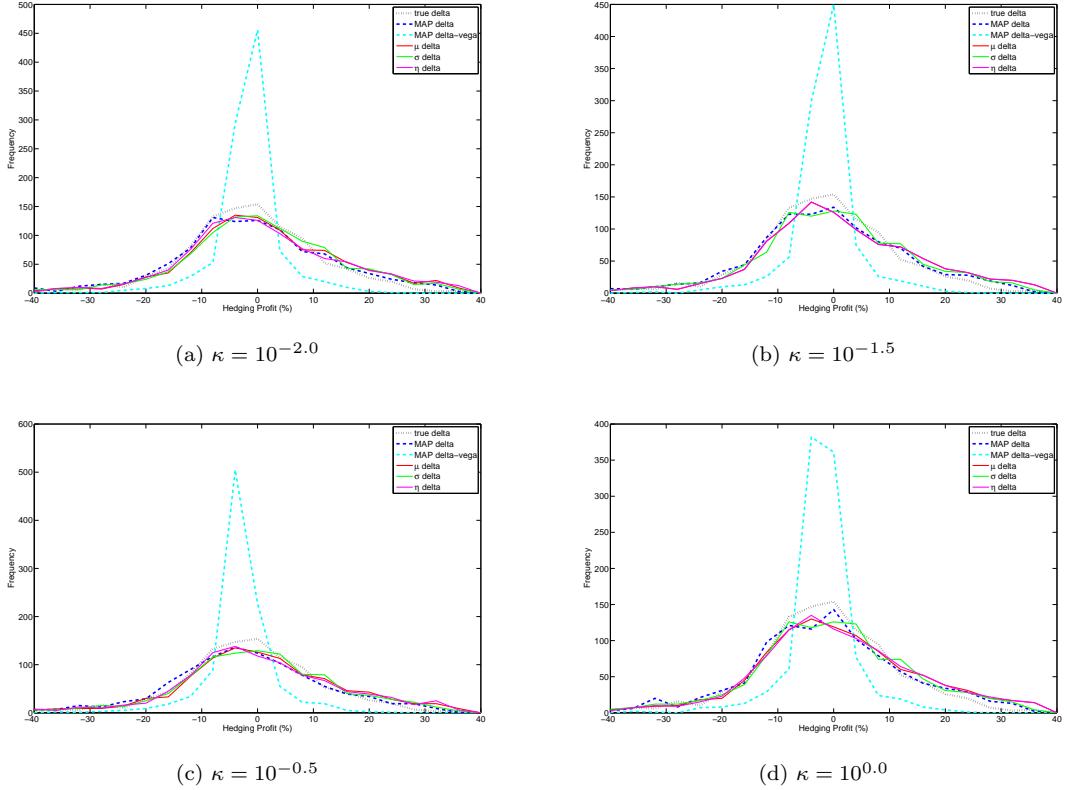
loss function $L(\theta, \theta')$ calculated by finding the losses incurred when simulating paths on surface θ and hedging throughout using surface θ' . But in our numerical experiment we have recalibrated the surfaces θ daily, and we have appropriately adjusted our MAP and Bayes hedges daily also. So the loss function is not perfectly synchronised with the actual hedging algorithm we implement. Hence, sometimes the changes do not fall within the quasi expected confidence intervals, and sometimes the intervals are very wide because some combinations (θ, θ') give very large losses and profits for the more sensitive barrier option. And these few large losses and profits stretch the variances of the absolute deviation and 5% shortfall hedges and hence widen the quasi predicted confidence interval. Nevertheless, at least for the mean hedge profit the quasi predicted interval is a good estimator for the realised improvement, and hence a useful a priori way of deciding whether or not to implement the Bayesian hedge.

4.4.6 Robustness

Next, just as we did in the previous two chapters, we shall repeat the hedging experiments for the samples found using different κ (form of the prior) and ε (observed data) in Section 2.5.6.

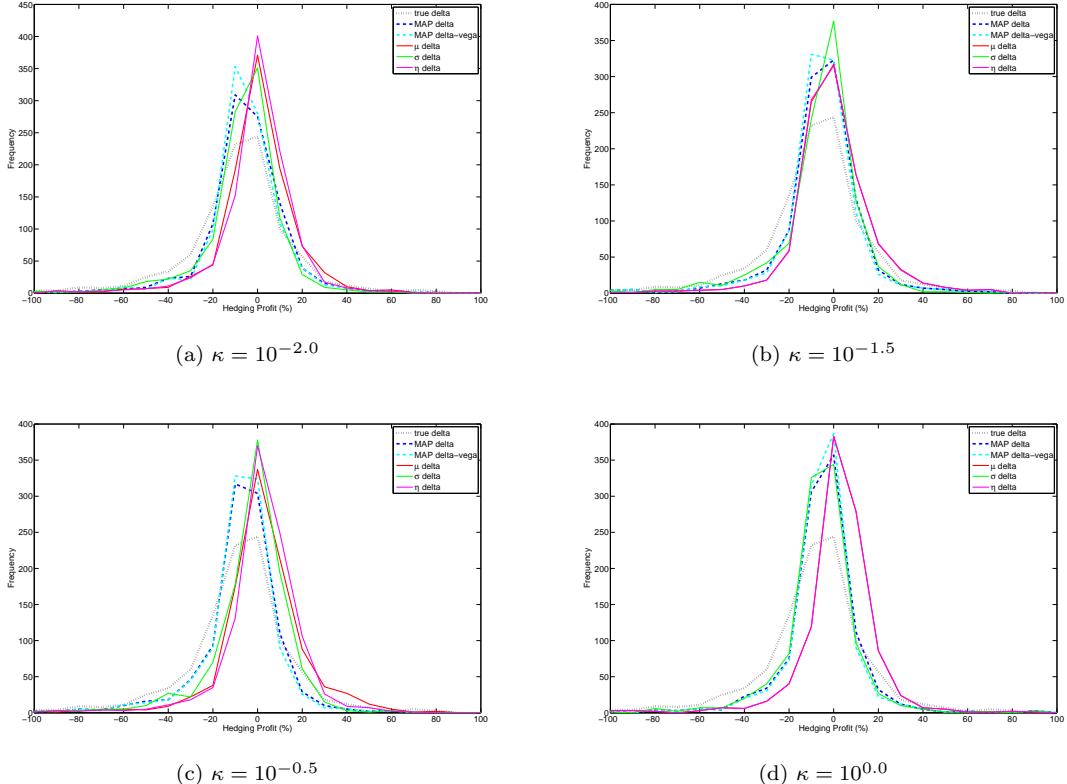
In Figure 4.7 and Figure 4.8 we see the effect on the hedging distributions of the call and barrier options for the first case (local volatility simulated paths) when varying the κ (the prior). Recall that we used $\kappa = 10^{-1.0}$ throughout the previous section's experiments. We observe that there is very little change in the performances of the hedging strategies for the call options. For the barrier option, however, there seems to be greater improvement as the value of κ increases to 1. For example, for $\kappa = 10^{-0.5}$ the hedging improvement of the Bayes hedges are 10.9% and 29.5% for the mean hedge profit and 5% profit shortfall respectively. However, with only four values of κ tested the pattern is inconclusive. But what we do see throughout is that the Bayes hedges consistently increase the mean hedge profit and 5% profit shortfall by at least 5% and 14% respectively. This suggests the Bayes method is robust to different values of κ .

In Figure 4.9 and Figure 4.10 we test robustness of the hedging improvements to noise added to the observed calibration prices. We use the results of Section 2.5.6 in which we added Gaussian noise with mean zero and standard deviation $\varepsilon = 10^{-3}$ of the original price to the prices. We choose the four out of the 100 experiments with $\varepsilon = 10^{-2.5}$ in which the MAP price of the option was furthest from the true price of the option, i.e. performed the worst. We label these four noises as A, B, C, D. Again, for the European call option we see very little change in the histograms of hedging profits and tabled statistics. The Bayes hedges consistently increase the mean hedge profit and 5% profit shortfall by between 0% and 3%. For the barrier option, however, we see that the MAP hedge can deteriorate hedge profits greatly. For example, for noise B, the MAP hedge's 5% profit shortfall falls to -127.7%, while the Bayes hedge's 5% profit shortfall remains around the -90% level. The Bayes hedges consistently increase mean relative hedge profits by over 9% and 5% profit shortfalls



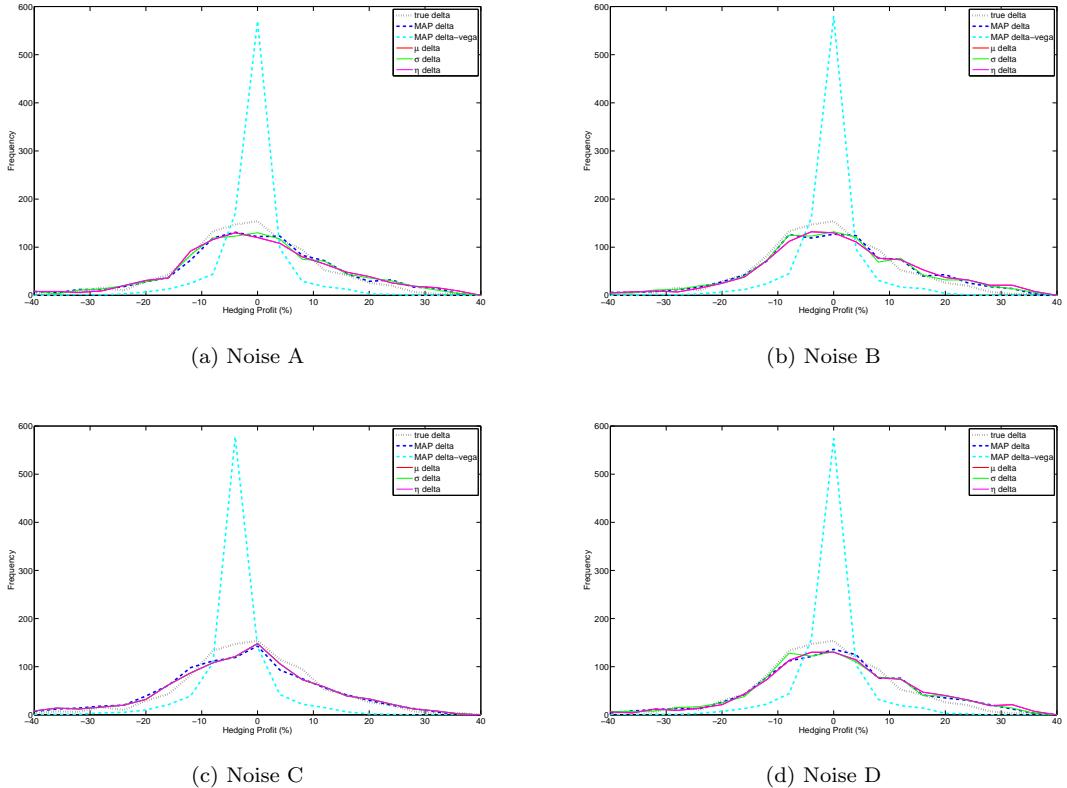
	strategy	mean hedge profit	absolute deviation	5% profit shortfall
$\kappa = 10^{-2.0}$ (600 surfaces calibrated)	true delta	0.7	12.3	-28.1
	MAP delta	1.3	14.9	-32.9
$\kappa = 10^{-1.5}$ (600 surfaces calibrated)	Bayes delta	3.5	14.6	-28.7
	improvement quasi predicted	+2.2 [1.8,1.8]	+0.3 [0.0,0.4]	+4.2 [2.4,3.5]
$\kappa = 10^{-0.5}$ (507 surfaces calibrated)	true delta	0.7	12.3	-28.1
	MAP delta	1.3	14.8	-32.8
$\kappa = 10^{0.0}$ (251 surfaces calibrated)	Bayes delta	3.6	14.7	-28.3
	improvement quasi predicted	+2.3 [1.9,1.9]	+0.1 [-0.1,0.4]	+4.5 [2.8,4.4]
$\kappa = 10^{-2.0}$ (600 surfaces calibrated)	true delta	0.7	12.3	-28.1
	MAP delta	0.0	14.9	-34.9
$\kappa = 10^{-1.5}$ (600 surfaces calibrated)	Bayes delta	3.0	14.7	-29.7
	improvement quasi predicted	+3.0 [2.5,2.6]	+0.3 [-0.0,0.6]	+5.2 [3.5,6.1]
$\kappa = 10^{-0.5}$ (507 surfaces calibrated)	true delta	0.7	12.3	-28.1
	MAP delta	0.0	14.9	-34.9
$\kappa = 10^{0.0}$ (251 surfaces calibrated)	Bayes delta	3.0	14.7	-29.7
	improvement quasi predicted	+1.9 [1.5,1.5]	+0.1 [-0.2,0.9]	+3.2 [1.5,2.7]

Figure 4.7: (upper) Histograms of hedging profits for European call option with strike 5000 ($S_0 = 5000$) and maturity 3 months for different prior parameter κ . (lower) Table of key statistics for each hedging strategy and the realised improvements and 95% quasi confidence intervals for the improvements of these statistics made by the Bayes hedges over the MAP delta hedge.



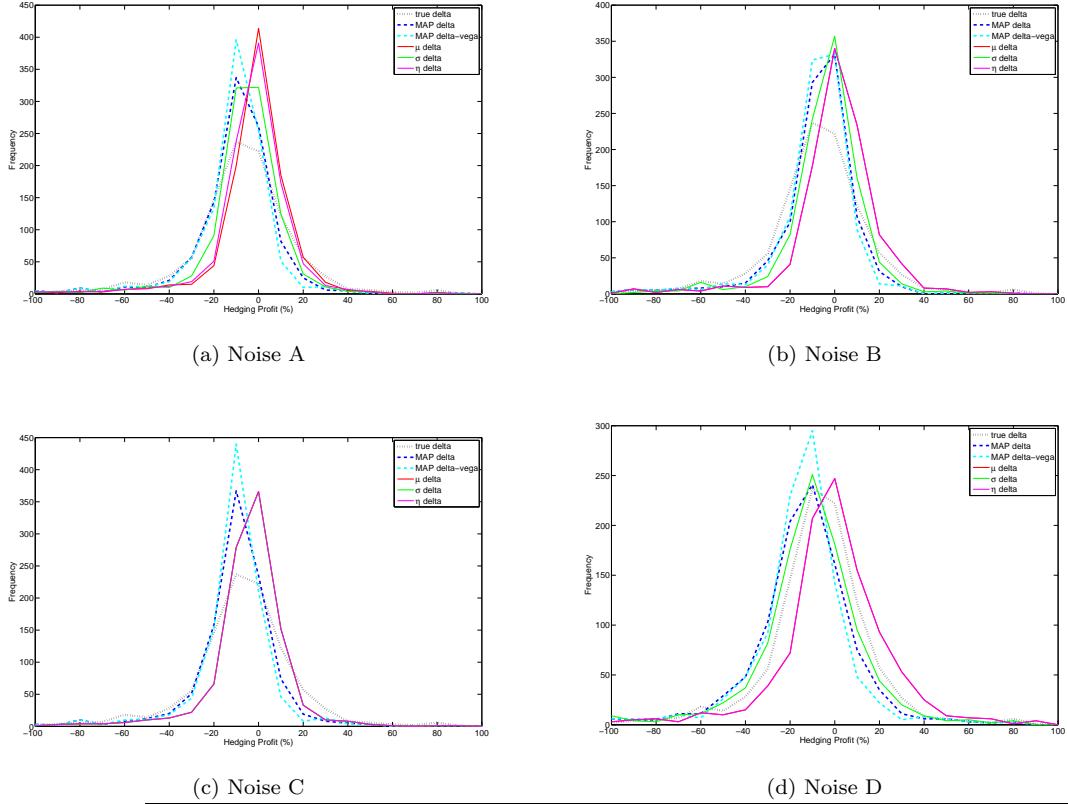
strategy	mean hedge profit	absolute deviation	5% profit shortfall
$\kappa = 10^{-2.0}$ (600 surfaces calibrated)	true delta	-6.5	46.2
	MAP delta	-3.5	33.5
	Bayes delta	2.3	34.2
	improvement quasi predicted	+5.8 [5.2,13.4]	-0.7 [-90.4,157.5]
$\kappa = 10^{-1.5}$ (600 surfaces calibrated)	true delta	-6.5	46.2
	MAP delta	-3.5	32.6
	Bayes delta	2.4	35.0
	improvement quasi predicted	+6.0 [6.1,13.0]	-2.1 [-81.4,137.6]
$\kappa = 10^{-0.5}$ (507 surfaces calibrated)	true delta	-6.5	46.2
	MAP delta	-5.9	34.7
	Bayes delta	4.9	32.3
	improvement quasi predicted	+10.9 [8.9,26.2]	+2.4 [-87.0,165.0]
$\kappa = 10^{0.0}$ (251 surfaces calibrated)	true delta	-6.5	46.2
	MAP delta	-2.6	29.6
	Bayes delta	5.6	30.1
	improvement quasi predicted	+8.2 [2.5,25.7]	-0.5 [-68.0,118.2]

Figure 4.8: (upper) Histograms of hedging profits for European call option with strike 5000 ($S_0 = 5000$), barrier 5500 and maturity 3 months for different prior parameter κ . (lower) Table of key statistics for each hedging strategy and the realised improvements and 95% quasi confidence intervals for the improvements of these statistics made by the Bayes hedges over the MAP delta hedge.



	strategy	mean hedge profit	absolute deviation	5% profit shortfall
noise A (32 surfaces calibrated)	<i>true delta</i>	0.7	12.3	-28.1
	<i>MAP delta</i>	2.1	14.7	-32.0
	<i>Bayes delta</i>	2.2	14.7	-30.8
	improvement quasi predicted	+0.1 [0.0,0.0]	-0.0 [-0.0,0.3]	+1.2 [0.0,0.0]
noise B (244 surfaces calibrated)	<i>true delta</i>	0.7	12.3	-28.1
	<i>MAP delta</i>	2.2	14.7	-31.8
	<i>Bayes delta</i>	3.4	14.8	-28.9
	improvement quasi predicted	+1.2 [0.9,1.0]	-0.0 [-0.2,0.3]	+2.9 [1.1,2.2]
noise C (2 surfaces calibrated)	<i>true delta</i>	0.7	12.3	-28.1
	<i>MAP delta</i>	-0.9	15.1	-37.3
	<i>Bayes delta</i>	-0.5	15.1	-37.1
	improvement quasi predicted	+0.4 [0.3,0.3]	0.0 [0.0,0.1]	+0.3 [0.4,0.5]
noise D (225 surfaces calibrated)	<i>true delta</i>	0.7	12.3	-28.1
	<i>MAP delta</i>	2.2	14.7	-31.9
	<i>Bayes delta</i>	3.2	14.7	-29.4
	improvement quasi predicted	+1.0 [0.8,0.8]	-0.0 [-0.1,0.3]	+2.5 [0.8,2.5]

Figure 4.9: (upper) Histograms of hedging profits for European call option with strike 5000 ($S_0 = 5000$) and maturity 3 months for different noises with mean $\varepsilon = 10^{-2.5}$. (lower) Table of key statistics for each hedging strategy and the realised improvements and 95% quasi confidence intervals for the improvements of these statistics made by the Bayes hedges over the MAP delta hedge.



	strategy	mean hedge profit	absolute deviation	5% profit shortfall
noise A (62 surfaces calibrated)	<i>true delta</i>	-5.5	46.1	-137.4
	<i>MAP delta</i>	-8.5	33.0	-115.9
	<i>Bayes delta</i>	1.0	30.4	-91.2
	improvement	+9.5	+2.6	+24.7
noise B (248 surfaces calibrated)	quasi predicted	[9.6,19.5]	[-101.0,211.9]	[15.6,134.7]
	<i>true delta</i>	-5.5	46.1	-137.4
	<i>MAP delta</i>	-7.3	35.4	-127.5
	<i>Bayes delta</i>	3.4	32.5	-90.2
	improvement	+10.7	+2.9	+37.3
noise C (13 surfaces calibrated)	quasi predicted	[10.7,23.0]	[-92.7,180.8]	[39.4,201.6]
	<i>true delta</i>	-5.5	46.1	-137.4
	<i>MAP delta</i>	-8.5	32.4	-112.6
	<i>Bayes delta</i>	-1.8	29.0	-94.1
	improvement	+10.1	+0.4	+12.6
noise D (72 surfaces calibrated)	quasi predicted	[3.8,6.5]	[-45.5,52.9]	[0.1,0.57.5]
	<i>true delta</i>	-5.5	46.1	-137.4
	<i>MAP delta</i>	-9.2	31.3	-105.5
	<i>Bayes delta</i>	0.2	30.2	-79.9
	improvement	+9.4	+1.1	+25.5
noise E (10 surfaces calibrated)	quasi predicted	[6.6,21.3]	[-135.4,284.4]	[-20.3,166.5]
	improvement	+10.1	+1.1	+25.5

Figure 4.10: (upper) Histograms of hedging profits for up-and-out barrier call option with strike 5000 ($S_0 = 5000$), barrier 5500 and maturity 3 months for different noises with mean $\varepsilon = 10^{-2.5}$. (lower) Table of key statistics for each hedging strategy and the realised improvements and 95% quasi confidence intervals for the improvements of these statistics made by the Bayes hedges over the MAP delta hedge.

by on average 20% which are close to the 7.2% and 22.1% improvements seen in Figure 4.4. Hence we again conclude robustness of the Bayes hedging improvements.

4.5 Extensions

The results of the previous two sections suggest the Bayesian hedges can be a very powerful way of increasing relative returns, especially for exotic options such as barrier options, when faced with model uncertainty. There are a variety of further avenues of research that naturally follow from these findings.

Firstly, given sufficient computing power, we could consider a dynamic loss function $L_t(\theta, \theta')$ rather than the static one we used. This will give more accurate Bayesian hedges, since we incorporate the newly observed calibration prices and update the loss function appropriately. Creating the loss matrix $L(\theta, \theta')$ can take over an hour, even using parallel computation. Though for an agent recalibrating and rehedging daily this is affordable at the start of the trading day, for practical testing using Monte Carlo this becomes unfeasible. For example, 1000 paths of 60 days rehedging for a 3-month option implies $1 \text{ hour} \times 1000 \times 60 = 6.8 \text{ years}$ to run.

The user is not restricted to delta hedging strategies. The methodology is easily applied to any hedging strategy where the hedge ratio is computed from the calibrated model. We could create a loss matrix $L(\theta, \theta')$ corresponding to delta-vega hedging or a delta-vega-gamma hedging. Indeed we can even extend the method to the mean-variance hedging introduced by Föllmer and Sondermann [68]. Recall that the strategy in mean-variance hedging is to minimise the variance V of the hedging errors given a pre-specified value of the mean hedging error E . In the Bayesian approach we would, for each model θ , find the expected minimum hedging error variance $V(\theta, \theta')$ when hedging using model θ' and fixed pre-specified mean hedging error E , and then set the loss function to $L(\theta, \theta') = V(\theta, \theta')$ or some non-decreasing function of $V(\theta, \theta')$.

It would also be worthwhile investigating whether tighter bounds could be found for the expected improvement. For example we might only consider the expectation (4.11) and variance (4.12) of a quantile of the improvements $I(\theta^0, \theta_L)$ rather than the full range. Also, instead of computing the first-order quasi confidence intervals, with sufficient power we can try to compute the actual confidence intervals. But, as explained, this is very demanding in terms of computation time as many paths need to be run on each surface and then the posterior recalibrated daily.

In light of the improvements seen by the three Bayesian hedges, an agent may choose to specify some other loss function that they wish to minimise. For example an agent might choose to minimise the greatest possible loss or maximise the average profit in the best 5% of cases. Furthermore, an agent may construct a loss function from a combination of other loss functions. For example,

analogous to mean-variance hedging, one might choose a loss function of the form

$$L^{g_\mu} + \alpha L^{g_\sigma}$$

for some risk-return tradeoff parameter α . This α might in turn be optimised in some sense, for example to minimise the expected loss given by (4.14).

Referring back to Section 4.3.3, we could also take traditional utility functions and transform them to loss functions using the relationship (4.19). For example, the exponential utility function given by $U(x) = -\exp\{-\gamma x\}$ for constant $\gamma > 0$ would give

$$g_{exp} = \exp\{-\gamma x\}$$

for constant $\gamma > 0$. And the power utility function $U(x) = x^\gamma/\gamma$ for $\gamma < 1, \gamma \neq 0$, gives

$$g_{pow} = -\frac{x^\gamma}{\gamma}$$

for $\gamma < 1, \gamma \neq 0$. It is easy to check that both g_{exp} and g_{pow} are decreasing and convex, so satisfy the assumptions of Theorem 4.3.3, and could thus also be used to construct penalty functionals for convex model uncertainty measures.

Chapter 5

Conclusions

In this thesis we have shown how to calibrate (non-parametric) financial models through a Bayesian framework. We have used the results of this calibration process to value the model uncertainty of different contracts and price and hedge these contracts. Throughout, the local volatility model was used as an example. We briefly summarise the findings of the previous three chapters in the next section, then interpret these findings and finally recommend some areas for future investigation in the third section.

5.1 Summary of Main Results

In Chapter 2 we considered the problem of marking a financial model to market before using it to price other contracts in order to prevent creating arbitrage opportunities. We described the ill-posedness of the problem caused by under-determinacy and market noise. Different regularisation techniques were discussed and then the Bayesian framework was introduced. It was observed that the Bayesian method can extract much more information than traditional regularised error function minimisation techniques which simply give an equivalent to the maximum a posteriori (MAP) model. Moreover, an equivalence between the two methods was observed in Remark 2.4.1. For the case of scalar volatility parameter, under some assumptions on the Bayesian prior, the market noise and pricing functions, we proved in Theorem 2.2.7 consistency of the Bayes estimates, i.e. that the Bayes estimator converges in probability to the true model parameter. We extended this result in Theorem 2.2.12 for finite-dimensional non-scalar volatility parameters, e.g. discretised local volatility surface. Finally, we applied our technique to two numerical experiments: one simulated using the local volatility example presented by Jackson, Suli & Howison [89], the other using real S&P 500 data presented in Coleman, Li & Verma [28]. In both cases we observed the variety of different shaped local volatility surfaces that could be fitted to the same prices. Bayesian prices for barrier and American options were then calculated using this distribution of surfaces and found to be closer than the MAP price to the true price. Sensitivity checks showed the MAP prices to be non-robust

to both changes in the Bayesian prior and market noise, compared with the Bayesian prices which were robust to both.

In Chapter 3 we introduced the concept of model uncertainty and contrasted it with market risk. Prompted by the model uncertainty measures introduced by Cont [30], we derived a wider class of coherent and convex model uncertainty measures. We adapted the axiomatic frameworks for market risk measures of Artzner et al. [2] and Frittelli and Gianin [69] to define new classes of coherent and convex model uncertainty measures respectively. In each case we found representation theorems akin to their market risk measure namesakes: Theorem 3.2.11 for coherent measures and Theorem 3.3.6 for convex measures. In each case we showed the Bayesian posterior distribution derived in Chapter 2 could be used with the representations to provide a rich class of model uncertainty measures. Lastly, we returned to the two numerical examples studied in Chapter 2 and found the convex model uncertainty values for European calls, barrier, and American options. Comments were made on the use of these values to rank the uncertainty of different products. Sensitivity checks were again carried out and it was concluded that the model uncertainty values were generally robust to deviations in the prior and market noise.

In Chapter 4, we extended the Bayesian approach further. In light of the model uncertainty quantified in the preceding chapter, Chapter 3, we devised optimal Bayesian hedging strategies. By optimal we mean that we specified loss functions, e.g. mean relative hedging loss and absolute deviation of relative hedging losses, and tried to find the hedging model which minimised the expected loss, where expectation was taken with respect to the Bayesian posterior found in Chapter 2. Depending on the accuracy of the posterior it was noted that the expected improvement (over the MAP model) was positive and quasi confidence intervals for the improvement could be estimated. The loss functions were used to construct more convex model uncertainty measures in Theorem 4.3.3 and a relationship was found between the loss functions and traditional utility functions. We then carried out two numerical experiments to test the performance of the optimal Bayesian delta hedges against MAP hedges. In the first case we generated paths using a local volatility surface and in the second case using a Heston model. For both experiments we hedged a European call option and an up-and-out barrier call option. In both cases we found the Bayesian hedges only marginally improved the hedging statistics for the European call. However, for the barrier option, the Bayesian hedges increased mean profits and shortfalls by up to 30%. Sensitivity analysis at the end showed these improvements to be robust to changes in the prior and market noise.

5.2 Interpretation of Findings

Chapter 2 gave three main contributions to the calibration problem. Firstly, we showed analytically that the Bayes estimator converges to the true value for an unknown finite-dimensional (but not necessarily scalar) parameter. Secondly, we devised a practical and fast way to implement the

Bayesian scheme. Thirdly, we showed using sensitivity checks that the Bayes estimator is robust. The interpretation of this is that the Bayesian method presented offers some practical advantages over traditional regularisation approaches in academia and industry. The crux of the Bayesian approach to calibration is to extract more information for very little extra effort. Especially in an age with high-speed parallel computing capabilities, the MCMC Metropolis routine can be optimised to be very fast. Moreover, this extra information is vital for producing stable and robust estimates for desired quantities such as prices of exotics, which is one of the primary goals for any option writer.

Following the variety of local volatility surfaces fitted to the same calibration prices in Chapter 2, Chapter 3 quantified the model uncertainty and found it to be significant, especially for exotic claims. The interpretation of this is clear: model uncertainty is highly prevalent, and option writers should be aware of it when deciding what premium to charge a client. We only numerically tested the case of model mis-estimation — when the model type is correct but the calibration parameter is unknown — and found large model uncertainty values. The model uncertainty attributable to model mis-specification — when not even the correct model type is known — is likely to be even larger, and agents need to be aware of this also. A claim that is found to have large model uncertainty values should be treated with caution and perhaps regulators or in-house risk monitors should set a threshold for model uncertainty values, beyond which an option is not issued or returned for better modelling and evaluation. And if this were to happen, then the axiomatic frameworks developed in Chapter 3 can be used to develop industry-standard model uncertainty measures for better comparison and regulation.

In light of the observed model uncertainty, Chapter 4 contributed practical methods for optimising hedging profits. The improvements were modest for the European call option but dramatic for the barrier call option. The interpretation of this is that option-issuing houses should in some way incorporate model uncertainty into their hedging strategies. Vega and Gamma hedges are a step in the right direction, but the hedge ratios are still in-model hedges and as such, susceptible to model mis-estimation or mis-specification. Again, finding the Bayesian hedge takes very little extra computational effort, but the rewards are significant. The Bayesian hedges do not necessarily cost more either, since we use the same strategy (delta or delta-vega or delta-vega-gamma) but simply optimise the hedge ratios with respect to the expected hedging error loss function. We even checked the robustness of these hedging strategies to variations in the Bayesian posterior by varying the form of the prior and increasing market noise. In both cases, the improvements were robust. The flexibility of the method is another attractive feature of the Bayesian hedging approach. That one can select the loss function according to one's preferences or financial motives is of huge importance to both those trying to maximise expected profits and those trying to minimise the variances of

hedging errors. Following explanation of the calibration method and quantification of the consequent model uncertainty, the optimal Bayesian hedges of Chapter 4 can be viewed as the greatest practical contribution of this thesis.

5.3 Recommendations for Future Research

At the end of each of the previous three chapters we discussed some extensions specific to the theme of the chapter. Here we discuss some alternative directions for the Bayesian approach to financial modelling.

One natural question that arises from the consistency theorems of Chapter 2 is whether we can identify when we are in the *incorrect* model. We showed that, under suitable conditions, convergence to the true parameter will occur when we assume the correct model type, but the converse has not yet been studied — whether convergence fails when we assume the wrong model type. Farmer [61] has shown Bayesian filtering to be able to distinguish between the perfect and imperfect model scenario in the case of the logistic map. A similar result in the context of financial models would be a very useful result to prove. On first thoughts it seems unlikely that this will hold. For example, if we assume a Black-Scholes model when the underlying actually follows a mean-reverting stochastic volatility model, such as the Heston model, then it is likely that the Bayesian estimate for the Black-Scholes volatility will converge to the long-term mean volatility. However, there may be an alternative indication of incorrect model choice — the rate of the convergence. We concentrated in Chapter 2 on whether consistency was achieved, but said nothing about the rate of the convergence. It is conceivable that convergence will be considerably slower when an incorrect model is chosen, and that this will then allow us to identify incorrect model choice.

A second question arising from the calibration problems we looked at is what is the information content of the calibration prices V . We used path-independent European calls to calibrate a local volatility model that we later used to price path-dependent options. However, perhaps using barrier options might improve the Bayesian posterior, in the sense of narrowing the distribution of the price of another path-dependent option. An investigation into how to pick the optimal number and type of calibration products, and how this optimality is affected by the products we later need to price, would be very useful.

It goes without saying that the Bayesian approach to calibration is very general and can be applied to any model where it is necessary to determine unobservable parameters. So, for example, we discussed in Example 2.6.1 exactly how we would construct the prior for the calibration of the jump diffusion model, and in particular the jump density function. But we could equally calibrate stochastic volatility models, general Lévy processes, and apply the Bayesian philosophy to liquidity and credit problems as well.

Although in Chapter 3 we reviewed market risk measures and then concentrated on constructing model uncertainty measures, there is scope for looking at combined measures. We can use the Bayesian posterior distribution of models to create a joint distribution on the product space $\Omega \times \Theta$, where recall Ω is the set of possible future scenarios and Θ is the set of possible models. For example, we could use a double expected shortfall measure. There might be a way of generalising market risk measures and model uncertainty measures to general measures on this product space. A joint consistent measure would be useful way to compare the different risk-uncertainty structures of varying financial products, and help an investor choose products based on their risk-uncertainty preferences.

Model uncertainty measures could also be used for determining the optimal number of parameters to use for a given model. Use very few parameters and the calibration problem is greatly under-determined, so many poor models are calibrated and the model uncertainty for a contract is high. Use very many parameters and the problem becomes over-determined, so too many very well calibrated but different models arise and the model uncertainty for a contract is again high. It is reasonable to hypothesise therefore that there is an optimum number of parameters which minimises the model uncertainty of products priced in that model. We should see a U-shaped curve if we plot model uncertainty versus number of parameters. This would be an excellent future line of research as knowing the optimum/correct number of parameters to use would be very useful to practitioners.

Similarly, we could use the model uncertainty values for different model types or classes of models to decide which model type or class to use. For example, suppose we model the underlying process in two ways: as a stochastic volatility process in which we try to calibrate the drift and diffusion coefficients, and also as a pure jump process in which we try to calibrate the jump density. Then, if the model uncertainty value for the stochastic volatility class of models is significantly lower than that for the class of pure jump models, we will be inclined to decide that the underlying is better modeled as having stochastic volatility than being a pure jump process. This application is referred to as *model selection* in the literature, and the papers by Raftery [115] and Wasserman [132] are good starting points. Using the model uncertainty value for model selection would be a very useful and important application in industry.

The hedging results presented in the previous chapter were very encouraging, and further computational studies would in themselves be a worthwhile venture. However, one might also try to apply the Bayesian philosophy to portfolio optimisation problems. In this setting the weights of assets in the portfolio would be the unknown parameters we are trying to determine. Again we would set our loss function to reflect the variance of the portfolio return or the expected loss of the portfolio or some other statistic of the portfolio return, and find the weights which minimise the expected value of the loss function. More work would need to be done to decide exactly how to construct the prior and likelihood functions, especially given there might be anywhere between 100 or 1000 assets

to choose weights for so the problem becomes very high-dimensional and the Metropolis sampling method we used is unlikely to still be computationally practical. However, there is a lot of scope for finding optimal Bayesian portfolio selections.

Appendix A

Datasets

A.1 Calibration Prices

Maturity	Strike (units of S_0)										
	0.80	0.90	0.94	0.96	0.98	1.00	1.02	1.04	1.06	1.10	1.20
0.083	1.003	0.507	0.312	0.219	0.136	0.070	0.030	0.010	0.003	0.000	0.000
0.167	1.010	0.518	0.332	0.246	0.168	0.104	0.058	0.029	0.012	0.001	0.000
0.250	1.012	0.531	0.352	0.270	0.196	0.132	0.083	0.048	0.025	0.004	0.000
0.500	1.029	0.577	0.414	0.337	0.265	0.200	0.146	0.102	0.068	0.024	0.000
0.750	1.052	0.623	0.469	0.396	0.327	0.264	0.208	0.160	0.119	0.059	0.004
1.000	1.079	0.671	0.525	0.457	0.390	0.329	0.274	0.224	0.180	0.110	0.021

Table A.1: For the simulated dataset: European call prices (units of 10^3) (using [89]) in Chapter 2 Section 2.5.

Maturity	Strike (units of S_0)									
	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20	1.30	1.40
0.175	91.3	62.8	35.2	12.9	2.1	0.1	0.0	0.0	0.0	0.0
0.425	96.3	69.0	44.0	23.3	8.5	2.3	0.4	0.2	0.0	0.0
0.695	101.8	76.1	52.6	32.6	16.4	5.9	1.9	0.6	0.1	0.0
0.940	106.8	82.2	59.9	39.9	23.8	11.3	4.7	1.8	0.2	0.0
1.000	108.0	83.6	61.6	41.6	25.4	12.8	5.5	2.1	0.2	0.1
1.500	117.2	94.4	73.1	54.0	37.3	23.7	14.3	7.7	1.9	0.3
2.000	125.7	104.0	83.6	64.9	48.2	34.2	23.6	14.7	5.6	1.8

Table A.2: For the S&P 500 dataset: European call prices (\$) (using [28]) in Chapter 2 Section 2.5.

A.2 Calibration Constants

Constant	Description	Test 1	Test 2
S_0	time 0 asset price	5000	\$590
r	rate of interest	0.05	0.0600
d	dividend yield	0.03	0.0262
σ_{atm}	time 0 ATM volatility	0.15	0.14
I	# of calibrating options	66	70
δ	calibration tolerance (b.p.)	3	4.5
M	# of nodes ($J \times L$)	27	32
λ_p	strength of prior (for sampling)	1	1
κ	Sobolev norm (LV model)	0.1	0.1
n	# of iterations in each MCMC chain	25000	25000
m	# of MCMC Metropolis chains	16	16
b	length of burn-in	1000	1000
k	frequency of thinning	100	100
du	jump function step size	0.00015	0.00013

Table A.3: Numerical examples constants for calibration process in Chapter 2.

Appendix B

Algorithms

B.1 Sobolev Norm Induced Inverse Covariance Matrix

Fitzpatrick [62] comments that the Sobolev norm can induce an inverse covariance matrix. Recall the definition for the functional $\|u\|_\kappa^2$ given by (2.24) in Chapter 2, Section 2.4.1. Let function $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ be twice differentiable in both variables and represented by the vector \tilde{u} corresponding to $M = J \times L$ nodes as described in Subsection 2.5.2. Then we can approximate $\|u\|_2^2$ by the quadrature rule,

$$\|u\|_\sim^2 = \tilde{u}^T \tilde{u} = \tilde{u}^T I \tilde{u},$$

where I is the $M \times M$ identity matrix. Consider the integral

$$\|\nabla u\|_2^2 = \int_S \int_t \left| \frac{\partial u}{\partial S} \right|^2 + \left| \frac{\partial u}{\partial t} \right|^2 dS dt$$

over the rectangle $[S_1, S_2] \times [t_1, t_2]$. Using the notation

$$\begin{aligned} u_{j,l} &= u(S_j, t_l), \\ \Delta_j^S &= S_{j+1} - S_j, \\ \Delta_l^t &= t_{l+1} - t_l, \end{aligned}$$

this integral can be approximated by

$$\begin{aligned} \|\nabla u\|_\sim^2 &= \left[\frac{1}{2} \left(\left| \frac{u_{2,1} - u_{1,1}}{\Delta_1^S} \right|^2 + \left| \frac{u_{2,2} - u_{1,2}}{\Delta_1^S} \right|^2 \right) + \frac{1}{2} \left(\left| \frac{u_{1,2} - u_{1,1}}{\Delta_1^t} \right|^2 + \left| \frac{u_{2,2} - u_{2,1}}{\Delta_1^t} \right|^2 \right) \right] \\ &\quad \times \Delta_1^S \Delta_1^t. \end{aligned}$$

Hence, if we represent the region $[S_{min}, S_{max}] \times [0, T_{max}]$ by the same J spatial points $S_{min} = s_1 < \dots < s_j < \dots < s_J = S_{max}$ and L temporal points $0 = t_1 < \dots < t_l < \dots < t_L$ chosen in

Subsection 2.5.2, then the approximation to the integral over the whole region becomes

$$\begin{aligned}\|\nabla u\|_{\sim}^2 &= \frac{1}{2} \sum_{j=1}^{J-1} \sum_{l=1}^{L-1} (u_{j+1,l} - u_{j,l}) \left[\frac{\Delta_l^t}{\Delta_j^S} 1_{l < L} + \frac{\Delta_{l-1}^t}{\Delta_j^S} 1_{l > 1} \right] \\ &+ \frac{1}{2} \sum_{j=1}^{J-1} \sum_{l=1}^{L-1} (u_{j,l+1} - u_{j,l}) \left[\frac{\Delta_j^S}{\Delta_l^t} 1_{j < J} + \frac{\Delta_{j-1}^S}{\Delta_l^t} 1_{j > 1} \right] \\ &= \tilde{u}^T Q \tilde{u},\end{aligned}\tag{B.1}$$

since (B.1) is a quadratic function of the elements of \tilde{u} and where Q is a semi-positive definite matrix. Writing

$$A^{-1} = \kappa I + Q$$

gives the result. Observe $\kappa I + Q$ is positive definite (provided $\kappa > 0$) so A exists.

Note that the norms and matrices could be calculated exactly for piecewise polynomials (i.e. the splines) but the extra computational effort outweighs the benefits of these calculations.

B.2 Fast Fourier Transform Formula for Pricing in the Heston Model

Recall the Heston stochastic volatility model [82] given in Chapter 4:

$$\begin{aligned}dS_t &= \mu S_t dt + \sigma_t S_t dW_t, \\ d[\sigma_t^2] &= \alpha(\beta - \sigma_t^2) + \gamma \sigma_t dZ_t,\end{aligned}$$

where $\mu, \alpha, \beta, \gamma$ are constants and ρ is the correlation between the Brownian motions W and Z and S_0 and σ_0^2 are the time 0 values of S and σ^2 respectively. The constant rate of riskless return will be denoted by r and the dividend rate by d .

To price an in-the-money or at-the-money European call option on S with strike K and maturity T , Carr & Madan [23] make the variable transformations $s = \log S_T$ and $k = \log K$. They then write the value at time 0 of the European call option as

$$C_T(k) \equiv \int_k^\infty e^{-rT} (e^s - e^k) q(s) ds,$$

where $q(s)$ is the risk-neutral density of the log-terminal price. By multiplying $C_T(k)$ by a dampening factor $\exp\{\lambda k\}$ for $\lambda > 0$ to make it square integrable and taking Fourier transforms, they arrive at the pricing formula

$$C_T(k) = e^{-\lambda k} \frac{1}{\pi} \int_0^\infty e^{-ivk} \psi(v) dv,\tag{B.2}$$

where

$$\psi(v) = \frac{e^{-rT} \phi(v - (\lambda + 1)i)}{\lambda^2 + \lambda - v^2 + i(2\lambda + 1)v}$$

and $\phi(u) = \mathbb{E}[\exp\{ius\}]$ is the characteristic function of $s = \log S_T$ in the risk-neutral measure. Heston [82] had earlier derived $\phi(u)$ as

$$\phi(u) = \exp\{C(T; u) + D(T; u) + iu \log S_0\},$$

where

$$\begin{aligned} C(T; u) &= iu(r - d)T + \frac{\alpha\beta}{\gamma^2} \left[(\alpha - iu\rho\gamma + h)T - 2 \log \left(\frac{1 - ge^{hT}}{1 - g} \right) \right], \\ D(T; u) &= \frac{\alpha - iu\rho\gamma + h}{\gamma^2} \left(\frac{1 - ge^{hT}}{1 - g} \right), \end{aligned}$$

and

$$\begin{aligned} g &= \frac{\alpha - iu\rho\gamma + h}{\alpha - iu\rho\gamma - h} \\ h &= \sqrt{(iu\rho\gamma - \alpha)^2 + \gamma^2(iu + u^2)}. \end{aligned}$$

Pricing out-of-the-money European call options requires slightly more manipulation because the Fourier inversion becomes highly oscillatory for very short maturities. Carr & Madan [23] provide a complimentary formula for out-of-the-money European calls, and the reader is referred to their paper for further details.

Carr & Madan approximate solution (B.2) using Simpson's quadrature by the following expression:

$$C_T(k_u) \approx e^{-\lambda k} \frac{1}{\pi} \sum_{j=0}^{j=N} e^{-i\frac{2\pi}{N}(j-1)(u-1)} e^{ibv_j} \psi(v_j) \frac{\eta}{3} (3 + (-1)^j - 1_{j=1}), \quad (\text{B.3})$$

where

$$\begin{aligned} k_u &= -b + \frac{2b}{N}(u-1) \text{ for } u = 1, \dots, N+1, \\ b &= \frac{\pi}{\eta}, \\ v_j &= \eta(j-1), \end{aligned}$$

N is the number of grid points and η is the grid coarseness. Hence, the approximation (B.3) simultaneously outputs the European call prices for all log strikes k_u for $u = 1, \dots, N+1$, i.e. log strikes in the interval $[-b, b]$. In their paper Carr & Madan [23] found the values $N = 4096$ and $\eta = 0.25$ to give a good compromise between speed and accuracy. We use the same values and take $\lambda = 1.25$.

Appendix C

Proofs

C.1 Dupire's Formula

We want to derive a volatility function σ for (2.10) that exactly reproduces the observed vanilla market prices $V(K, T)$, where $V(K, T)$ is the price at time 0 of a European Call option on S_T with strike K and maturity T . Formally, we assume that through an arbitrage-free model of the stock price we have the entire family $V(K, T)$ of European Call options prices, and that there is a probability space $(\Omega, \mathcal{F}, \mathbb{P}^*)$ on which there exists a martingale M for which

$$V(K, T) = \mathbb{E}_{\mathbb{P}^*} (M_T - Ke^{-rT})^+. \quad (\text{C.1})$$

Recall that a function $\sigma : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}$ is called the local volatility if prices of all European Call options under the model (2.10) agree with a given family of prices $V(K, T)$, i.e. $V(K, T) = \mathbb{E}_{\mathbb{P}^*}[B_T^{-1}(S_T - K)^+]$ for all values of T and K , where $B_T = e^{-rT}$

We now proceed to derive Dupire's PDE for the local volatility function σ and an expression for it in terms of the observed call prices $V(K, T)$ defined above, in the absence of arbitrage. Suppose that the local volatility function σ is such that (2.10) has unique solution S and for any $t \in \mathbb{R}^+$ the random variable S_t has a continuous density function $f(s, t)$. Let $V(K, T)$ be as defined above and assume $V \in C^{2,1}$. Then we show that V satisfies the Dupire's PDE

$$-V_T(K, T) + \frac{1}{2}\sigma^2(K, T)K^2V_{KK}(K, T) - rKV_K(K, T) = 0, \quad (\text{C.2})$$

with the initial condition $V(K, 0) = (S_0 - K)^+$.

We reproduce Dupire's original proof found in Musiela and Rutkowski [110]. Let $B_t = e^{rt}$ and take $\tilde{S}_t = e^{-rt}S_t$ and the density function $f(s, t)$ of S_t . Then (2.10) becomes

$$d\tilde{S}_t = \sigma(S_t, t)\tilde{S}_tdW_t$$

under measure \mathbb{P}^* , and hence \tilde{S} is a continuous local martingale. Under further assumptions [110]

we can show \tilde{S} is a martingale and so, by the definition given in (C.1),

$$\begin{aligned} V(K, T) &= \mathbb{E}_{\mathbb{P}^*}(\tilde{S}_T - e^{-rt}K)^+ \\ &= e^{-rt}\mathbb{E}_{\mathbb{P}^*}[(S_T - K)^+] \\ &= e^{-rt} \int_K^\infty (s - K)f(s, T) ds. \end{aligned}$$

Differentiating the above twice with respect to K gives

$$V_{KK}(K, T) = e^{-rt}f(K, T).$$

We divide through by e^{-rt} and differentiate once with respect to T using the product rule to get

$$re^{-rt}V_{KK}(K, T) + e^{-rt}V_{KKT}(K, T) = \frac{\partial f}{\partial T}(K, T).$$

Under suitable regularity conditions, we [94] that the density function f also satisfies the Fokker-Planck (or forward Kolmogorov) equation:

$$-\frac{\partial f}{\partial T}(K, T) + \frac{1}{2} \frac{\partial^2}{\partial K^2}(K^2\sigma^2(K, T)f(K, T)) - \frac{\partial}{\partial K}(rKf(K, T)) = 0.$$

Substituting into this the expressions for $f(K, t)$ and $\frac{\partial f}{\partial t}(K, t)$ we found above and then dividing throughout by e^{rT} , we get

$$\begin{aligned} 0 &= -rV_{KK} - V_{KKT} - rKV_{KKK} - rV_{KK} + \frac{1}{2} \frac{\partial^2}{\partial K^2}(K^2\sigma^2V_{KK}) \\ &= -2rV_{KK} - V_{KKT} - rKV_{KKK} + \frac{1}{2} \frac{\partial^2}{\partial K^2}(K^2\sigma^2V_{KK}) \\ &= \frac{\partial^2}{\partial K^2} \{-rKV_K - V_T + \frac{1}{2}K^2\sigma^2V_{KK}\}. \end{aligned}$$

Integrating twice with respect to K and using again that $c_{KK}(K, T) = e^{-rT}f(K, T)$, we get

$$-rKV_K - V_T + \frac{1}{2}K^2\sigma^2V_{KK} = \varepsilon(T)K + \eta(T)$$

for some functions $\varepsilon(T)$ and $\eta(T)$. Since the partial derivatives of V tend to zero as $K \rightarrow \infty$ for all T , we have that the $\lim_{K \rightarrow \infty} \varepsilon(T)K + \eta(T) = 0$ for all T and hence that $\varepsilon = \eta = 0$. Thus we get Dupire's PDE (C.2). Rearranging Dupire's PDE (C.2) gives Dupire's formula (2.15).

C.2 Terminal Hedging Error for Discrete Delta-Neutral Portfolio

We show by induction that

$$\begin{aligned} &\Pi_{t_k}(S(\theta), \Delta(\theta')) - \Delta_{t_k}(\theta')S_{t_k}(\theta) \\ &= f_0^X(\theta')e^{rt_k} + \sum_{i=0}^k [\Delta_{t_{i-1}}(\theta')(1 + q(t_i - t_{i-1})) - \Delta_{t_i}(\theta')]S_{t_i}(\theta)e^{r(t_k - t_i)} \end{aligned} \quad (\text{C.3})$$

for $k = 0, \dots, n$. We suppress the arguments $\theta, \theta', S, \Delta$ and write $\delta_i = t_i - t_{i-1}$ for brevity. For $k = 0$ the value of the delta hedging portfolio is given by

$$\Pi_{t_0} - \Delta_{t_0} S_{t_0} = f_{t_0}^X - \Delta_{t_0} S_{t_0},$$

which is of the form of (C.3). Next for k we have

$$\begin{aligned} & \Pi_{t_k} - \Delta_{t_k} S_{t_k} \\ &= \Delta_{t_{k-1}} S_{t_k} + (\Pi_{t_{k-1}} - \Delta_{t_{k-1}} S_{t_{k-1}}) e^{r\delta_k} + q\delta_k \Delta_{t_{k-1}} S_{t_k} - \Delta_{t_k} S_{t_k} \\ &= (\Pi_{t_{k-1}} - \Delta_{t_{k-1}} S_{t_{k-1}}) e^{r\delta_k} + [\Delta_{t_{k-1}}(1 + q\delta_k) - \Delta_{t_k}] S_{t_k}, \end{aligned}$$

which is of the form of (C.3). Setting $k = n$ gives (4.22).

C.3 Terminal Hedging Error for Discrete Delta-Vega-Neutral Portfolio

We show by induction that

$$\begin{aligned} & \Pi_{t_k}(S(\theta), \Delta(\theta')) - \Delta_{t_k}(\theta') S_{t_k}(\theta) - \Xi_{t_k}(\theta') f_{t_k}^Y(\theta) \\ &= f_0^X(\theta') e^{rt_k} + \sum_{i=0}^k [\Delta_{t_{i-1}}(\theta')(1 + q(t_i - t_{i-1})) - \Delta_{t_i}(\theta')] S_{t_i}(\theta) e^{r(t_k - t_i)} \end{aligned} \quad (\text{C.4})$$

$$+ \sum_{i=0}^k [\Xi_{t_{i-1}}(\theta') - \Xi_{t_i}(\theta')] f_{t_i}^Y(\theta) e^{r(t_k - t_i)} \quad (\text{C.5})$$

(C.6)

for $k = 0, \dots, n$. We suppress the arguments $\theta, \theta', S, \Delta$ and write $\delta_i = t_i - t_{i-1}$ for brevity. For $k = 0$ the value of the delta-vega hedging portfolio is given by

$$\Pi_{t_0} - \Delta_{t_0} S_{t_0} - \Xi_{t_0} f_{t_0}^Y = f_{t_0}^X - \Delta_{t_0} S_{t_0} - \Xi_{t_0} f_{t_0}^Y,$$

which is of the form of (C.5). Next, for general k we have

$$\begin{aligned} & \Pi_{t_k} - \Delta_{t_k} S_{t_k} - \Xi_{t_k} f_{t_k}^Y \\ &= \Delta_{t_{k-1}} S_{t_k} + -\Xi_{t_{k-1}} f_{t_k}^Y (\Pi_{t_{k-1}} - \Delta_{t_{k-1}} S_{t_{k-1}} - \Xi_{t_{k-1}} f_{t_{k-1}}^Y) e^{r\delta_k} \\ &+ d\delta_k \Delta_{t_{k-1}} S_{t_k} - \Delta_{t_k} S_{t_k} - \Xi_{t_k} f_{t_k}^Y \\ &= (\Pi_{t_{k-1}} - \Delta_{t_{k-1}} S_{t_{k-1}} - \Xi_{t_{k-1}} f_{t_{k-1}}^Y) e^{r\delta_k} \\ &+ [\Delta_{t_{k-1}}(1 + q\delta_k) - \Delta_{t_k}] S_{t_k} + [\Xi_{t_{k-1}} - \Xi_{t_k}] f_{t_k}^Y, \end{aligned}$$

which is of the form of (C.5). Setting $k = n$ gives (4.22).

References

- [1] A. Apte, M. Hairer, A.M. Stuart, and J. Voss. Sampling the posterior: An approach to non-Gaussian data assimilation. *Physica D: Nonlinear Phenomena*, 230(1-2):50–64, 2007.
- [2] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. *Risk management: value at risk and beyond*, page 145, 2002.
- [3] R.C. Aster, B. Borchers, and C.H. Thurber. *Parameter estimation and inverse problems*. Academic Press, 2005.
- [4] M. Avellaneda, C. Friedman, R. Holmes, and D. Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance*, 4:37–64, 1997.
- [5] M. Avellaneda, A. Lévy, and A. Paras. Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance*, 2:73–88, 1995.
- [6] M. Avellaneda and A. Paras. Managing the volatility risk of portfolios of derivative securities: the Lagrangian uncertain volatility model. *Applied Mathematical Finance*, 3(1):21–52, 1996.
- [7] A. Barron, M.J. Schervish, and L. Wasserman. The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics*, 27(2):536–561, 1999.
- [8] A.K. Basu. *Measure theory and probability*. PHI Learning Pvt. Ltd., 1999.
- [9] D. Belomestny and M. Reiß. Spectral calibration of exponential Lévy models. *Finance and Stochastics*, 10(4):449–474, 2006.
- [10] H. Berestycki, J. Busca, and I. Florent. Asymptotics and calibration of local volatility models. *Quantitative Finance*, 2(1):61–69, 2002.
- [11] J.O. Berger. *Statistical decision theory and Bayesian analysis*. Springer, 1985.
- [12] A. Beskos and A. Stuart. MCMC methods for sampling function space. In *Sixth International Congress on Industrial and Applied Mathematics: Zurich, Switzerland, July 16-20, 2007*, page 337. European Mathematical Society, 2009.

- [13] R. Bhar, C. Chiarella, H. Hung, and W.J. Runggaldier. The volatility of the instantaneous spot interest rate implied by arbitrage pricing: A dynamic bayesian approach. *Automatica*, 42(8):1381–1393, 2006.
- [14] P.J. Bickel and J.A. Yahav. Some contributions to the asymptotic theory of Bayes solutions. *Probability Theory and Related Fields*, 11(4):257–276, 1969.
- [15] N.H. Bingham and R. Kiesel. Semi-parametric modelling in finance: theoretical foundations. *Quantitative Finance*, 2(4):241–250, 2002.
- [16] N.H. Bingham and R. Kiesel. *Risk-neutral valuation: Pricing and hedging of financial derivatives*. Springer Verlag, 2004.
- [17] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637, 1973.
- [18] J.N. Bodurtha and M. Jermakyan. Non-Parametric Estimation of an Implied Volatility Surface. *Journal of Computational Finance*, 2:29–60, 1999.
- [19] N. Branger and C. Schlag. Can tests based on option hedging errors correctly identify volatility risk premia. *Goethe University Department of Finance Working Paper*, 136, 2004.
- [20] N. Branger and C. Schlag. Model Risk: A Conceptual Framework for Risk Measurement and Hedging. Technical report, Working paper, Goethe University, 2004.
- [21] M. Britten-Jones and A. Neuberger. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 55(2):839–866, 2000.
- [22] P. Carr, H. Geman, D.B. Madan, and M. Yor. The Fine Structure of Asset Returns: An Empirical Investigation. *The Journal of Business*, 75(2):305–332, 2002.
- [23] P. Carr and D. Madan. Option valuation using the fast fourier transform. *Journal of Computational Finance*, 2(4):61?73, 1999.
- [24] P. Carr and D. Madan. Determining volatility surfaces and option values from an implied volatility smile. In *Quantitative analysis in financial markets: collected papers of the New York University Mathematical Finance Seminar*, page 163. World Scientific Pub Co Inc, 2001.
- [25] P. Carr and L. Wu. Static Hedging of Standard Options. *CRIF Working Paper series*, page 22, 2002.
- [26] C. Chiarella, M. Craddock, and N. El-Hassan. The calibration of stock option pricing models using inverse problem methodology. *QFRQ Research Papers, UTS Sydney*, 2000.

- [27] T.F. Coleman, Y. Kim, Y. Li, and A. Verma. Dynamic Hedging With a Deterministic Local Volatility Function Model. 2003.
- [28] T.F. Coleman, Y. Li, and A. Verma. Reconstructing the unknown local volatility function. In *Quantitative analysis in financial markets: collected papers of the New York University Mathematical Finance Seminar*, page 192, 2001.
- [29] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.
- [30] R. Cont. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3):519–547, 2006.
- [31] R. Cont and P. Tankov. Calibration of jump-diffusion option pricing models: a robust non-parametric approach. 2002.
- [32] R. Cont and P. Tankov. *Financial modelling with jump processes*. CRC Pr I Llc, 2004.
- [33] R. Cont and P. Tankov. Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem. *SIAM Journal on Control and Optimization*, 45(1):1–25, 2006.
- [34] R. Cont, P. Tankov, and E. Voltchkova. Hedging with options in models with jumps. In *Proceedings of the 2005 Abel Symposium in Honor of Kiyosi Ito*, 2005.
- [35] P. Contreras and S.E. Satchell. A Bayesian Confidence Interval for Value-at-Risk. *Cambridge Working Papers in Economics*, 2003.
- [36] J.C. Cox, J.E. Ingersoll Jr, and S.A. Ross. A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, pages 385–407, 1985.
- [37] S. Crepey. Calibration of the local volatility in a generalized Black-Scholes model using Tikhonov regularization. *SIAM Journal on Mathematical Analysis*, 34(5):1183–1206, 2003.
- [38] S. Crepey. Calibration of the local volatility in a trinomial tree using Tikhonov regularization. *Inverse Problems*, 19(1):91–127, 2003.
- [39] S. Crepey. Delta-hedging vega risk? *Quantitative Finance*, 4(5):559–579, 2004.
- [40] T. Darsinos and S. Satchell. Bayesian analysis of the Black-Scholes option price. *Forecasting expected returns in the financial markets*, page 117, 2007.
- [41] T. Darsinos and S. Satchell. Bayesian forecasting of options prices: natural framework for pooling historical and implied volatility information1. *Forecasting expected returns in the financial markets*, page 151, 2007.

- [42] M. Davis. Mathematics of financial markets. *Mathematics Unlimited*, 2001.
- [43] M.H.A. Davis. Complete-market models of stochastic volatility. *Proceedings: Mathematics, Physical and Engineering Sciences*, pages 11–26, 2004.
- [44] E. Derman. Model risk. *Risk*, 9(5):34–37, 1996.
- [45] E. Derman, I. Kani, and N. Chriss. Implied Trinomial Trees of the Volatility Smile. *Journal of Derivatives*, 4:7–22, 1996.
- [46] E. Derman, I. Kani, and J.Z. Zou. The local volatility surface: Unlocking the information in index option prices. *Financial Analysts Journal*, 52(4):25–36, 1996.
- [47] J. Detemple and C. Osakwe. The valuation of volatility options. *Review of Finance*, 4(1):21, 2000.
- [48] P. Diaconis and D. Freedman. On the consistency of Bayes estimates. *The Annals of Statistics*, pages 1–26, 1986.
- [49] J.L. Doob. Application of the theory of martingales. *Le calcul des probabilités et ses applications*, pages 23–27, 1949.
- [50] B. Dumas, J. Fleming, and R.E. Whaley. Implied volatility functions: Empirical tests. *The Journal of Finance*, 53(6):2059–2106, 1998.
- [51] B. Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.
- [52] H. Egger and H.W. Engl. Tikhonov regularization applied to the inverse problem of option pricing: convergence analysis and rates. *Inverse Problems*, 21(3):1027–1045, 2005.
- [53] H. Egger, T. Hein, and B. Hofmann. On decoupling of volatility smile and term structure in inverse option pricing. *Inverse Problems*, 22:1247–1259, 2006.
- [54] N. El Karoui, M. Jeanblanc-Picque, and S.E. Shreve. Robustness of the Black and Scholes formula. *Mathematical Finance*, 8:93–126, 1998.
- [55] J. Elder. Hedging for Financial Derivatives. *University of Oxford, Ph. D. Thesis*, 2002.
- [56] M. Elliott. Controlling model risk. *Derivatives Strategy*, 1997.
- [57] D. Ellsberg. Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, pages 643–669, 1961.
- [58] H.W. Engl. Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Journal of Optimization Theory and Applications*, 52(2):209–215, 1987.

- [59] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Kluwer Academic Pub, 1996.
- [60] H.W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularization of nonlinear ill-posed problems. *Inverse Problems*, 5(4):523–540, 1989.
- [61] C.L. Farmer. Bayesian statistics, inverse problems and the logistic map. *Submitted*, 2007.
- [62] B.G. Fitzpatrick. Bayesian analysis in inverse problems. *Inverse problems*, 7(5):675–702, 1991.
- [63] H. Föllmer and P. Leukert. Quantile hedging. *Finance and Stochastics*, 3(3):251–273, 1999.
- [64] H. Föllmer and P. Leukert. Efficient hedging: cost versus shortfall risk. *Finance and Stochastics*, 4(2):117–146, 2000.
- [65] H. Föllmer and I. Penner. Convex risk measures and the dynamics of their penalty functions. *Statistics & Decisions*, 24(1/2006):61–96, 2006.
- [66] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, 2002.
- [67] H. Föllmer and A. Schied. Stochastic finance: an introduction in discrete time. 2004.
- [68] H. Föllmer and D. Sondermann. Hedging of non-redundant contingent claims. 1986.
- [69] M. Frittelli and E.R. Gianin. Putting order in risk measures. *Journal of Banking and Finance*, 26(7):1473–1486, 2002.
- [70] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, 2nd edition, 2004.
- [71] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472, 1992.
- [72] S. Ghosal. A review of consistency and convergence rates of posterior distributions. In *Proc. Varanasi Symp. on Bayesian Inference*, 1998.
- [73] S. Ghosal, J.K. Ghosh, and A.W. van der Vaart. Convergence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531, 2000.
- [74] P. Glasserman. *Monte Carlo methods in financial engineering*. Springer Verlag, 2003.
- [75] G.H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

- [76] J. Gondzio, R. Kouwenberg, and T. Vorst. Hedging options under transaction costs and stochastic volatility. *Journal of Economic Dynamics and Control*, 27(6):1045–1068, 2003.
- [77] T.C. Green and S. Figlewski. Market risk and model risk for a financial institution writing options. *Journal of Finance*, pages 1465–1499, 1999.
- [78] A. Grünbichler and F.A. Longstaff. Valuing futures and options on volatility. *Journal of Banking and Finance*, 20(6):985–1001, 1996.
- [79] S.B. Hamida and R. Cont. Recovering volatility from option prices by evolutionary optimization. *Journal of Computational Finance*, 8(4), 2005.
- [80] P.C. Hansen. *The L-curve and its use in the numerical treatment of inverse problems*. IMM, Department of Mathematical Modelling, Technical University of Denmark, 2001.
- [81] J.M. Henderson and R.E. Quandt. *Microeconomic theory: A mathematical approach*. McGraw-Hill New York, 1958.
- [82] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, page 327?343, 1993.
- [83] B. Hilberink and L.C.G. Rogers. Optimal capital structure and endogenous default. *Finance and Stochastics*, 6(2):237–263, 2002.
- [84] P. Huber. *Robust Statistics*. Wiley, 1981.
- [85] J. Hull. *Options, futures, and other derivatives*. Pearson Prentice Hall, 2008.
- [86] J. Hull and W. Suo. A methodology for assessing model risk and its application to the implied volatility function model. *The Journal of Financial and Quantitative Analysis*, 37(2):297–318, 2002.
- [87] J. Hull and A. White. Hedging the risks from writing foreign currency options. *Journal of International money and Finance*, 6(2):131–152, 1987.
- [88] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *Journal of finance*, pages 281–300, 1987.
- [89] N. Jackson, E. Suli, and S. Howison. Computation of Deterministic Volatility Surfaces. *Journal of Computational Finance*, 2(2):5–32, 1999.
- [90] E. Jacquier and R. Jarrow. Bayesian analysis of contingent claim model error. *Journal of Econometrics*, 94(1-2):145–180, 2000.

- [91] E. Jacquier, N.G. Polson, and P.E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 12(4):371–389, 1994.
- [92] A. Jobert, A. Platania, and L.C.G. Rogers. A Bayesian solution to the equity premium puzzle. *Statistical Laboratory, University of Cambridge*, 2006.
- [93] M. Jonsson and R. Sircar. Partial hedging in a stochastic volatility environment. *Mathematical Finance*, 12:375–409, 2002.
- [94] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus*. Springer, 1991.
- [95] J. Kerkhof and B. Melenberg. Backtesting for risk-based regulatory capital. *Journal of Banking and Finance*, 28(8):1845–1865, 2004.
- [96] J. Kerkhof, B. Melenberg, and H. Schumacher. Model risk and regulatory capital. *Discussion Papers*, 12, 2006.
- [97] F.H. Knight. Risk, Uncertainty and Profit, 1921.
- [98] R. Lagnado and S. Osher. A technique for calibrating derivative security pricing models: numerical solution of an inverse problem. *Journal of computational finance*, 1(1):13–25, 1997.
- [99] L. Le Cam. On some asymptotic properties of maximum likelihood estimates and related Baye’s estimates. *Univ. California Publ. Stat.*, 1:277–330, 1953.
- [100] E.L. Lehmann and G. Casella. *Theory of point estimation*. Springer New York, 1998.
- [101] A. Li. Model Calibration, Risk Measurement, and the Hedging of Derivatives. *Barclays Capital Working Paper Series*, 1999.
- [102] T.J. Lyons. Uncertain volatility and the risk-free synthesis of derivatives. *Applied Mathematical Finance*, 2:117–133, 1995.
- [103] D.W. Marquardt. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612, 1970.
- [104] K. Matusita. On the estimation by the minimum distance method. *Annals of the Institute of Statistical Mathematics*, 5(1):59–65, 1953.
- [105] K. Matusita. Decision rules, based on the distance, for problems of fit, two samples, and estimation. *The Annals of Mathematical Statistics*, pages 631–640, 1955.
- [106] M.L. McIntyre. Empirical Tests of an Option Price Inversion Approach. *Rotman School of Management-Finance*, 1999.

- [107] R.C. Merton. On the pricing of corporate debt: The risk structure of interest rates. *Journal of finance*, pages 449–470, 1974.
- [108] M. Monoyios. Optimal hedging and parameter uncertainty. *IMA Journal of Management Mathematics*, 18(4):331, 2007.
- [109] V.A. Morozov. *Methods for solving incorrectly posed problems*. Springer, 1984.
- [110] M. Musiela and M. Rutkowski. *Martingale methods in financial modelling*. Springer Verlag, 2005.
- [111] Basel Committe on Banking Supervision. Basel ii: Third consultative paper. *Bank for International Settlements*, 1999.
- [112] Basel Committe on Banking Supervision. A new capital adequacy framework. *Bank for International Settlements*, 1999.
- [113] H. Para and C. Reisinger. Calibration of instantaneous forward rate volatility in a bayesian framework. (*working paper*), 2007.
- [114] D. Psychoyios and G. Skiadopoulos. Volatility options: Hedging effectiveness, pricing, and model error. *Journal of Futures Markets*, 26(1):1–32, 2006.
- [115] A.E. Raftery. Bayesian model selection in structural equation models. *Testing structural equation models*, pages 163–180, 1993.
- [116] R. Rebonato. *Volatility and correlation: the perfect hedger and the fox*. Wiley, 2004.
- [117] T. Roths, M. Marth, J. Weese, and J. Honerkamp. A generalized regularization method for nonlinear ill-posed problems enhanced for nonlinear regularization terms. *Computer physics communications*, 139(3):279–296, 2001.
- [118] M. Rubinstein. Implied binomial trees. *Journal of Finance*, pages 771–818, 1994.
- [119] D. Samperi. Calibrating a diffusion pricing model with uncertain volatility: regularization and stability. *Mathematical Finance*, 12(1):71–87, 2002.
- [120] O. Scherzer. The use of Morozov’s discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems. *Computing*, 51(1):45–60, 1993.
- [121] O. Scherzer, H.W. Engl, and K. Kunisch. Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. *SIAM Journal on Numerical Analysis*, 30(6):1796–1838, 1993.
- [122] L. Schwartz. On Bayes procedures. *Probability Theory and Related Fields*, 4(1):10–26, 1965.

- [123] L. Scott. Random variance option pricing. *Advances in Futures and Options Research*, 5:113–135, 1991.
- [124] X. Shen and L. Wasserman. Rates of convergence of posterior distributions. *The Annals of Statistics*, 29(3):687–714, 2001.
- [125] N. Shephard. Statistical aspects of ARCH and stochastic volatility. *Time series models in econometrics, finance and other fields*, 65:1–67, 1996.
- [126] D.S. Sivia. *Data analysis: a Bayesian tutorial*. Oxford University Press, USA, 2006.
- [127] E.M. Stein and J.C. Stein. Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies*, pages 727–752, 1991.
- [128] H. Strasser. Consistency of maximum likelihood and Bayes estimates. *The Annals of Statistics*, 9(5):1107–1113, 1981.
- [129] P. Tankov. Lévy processes in finance and risk management. *Wilmott Magazine, September-October*, 2007.
- [130] A.N. Tikhonov, V.Y. Arsenin, and F. John. *Solutions of ill-posed problems*. VH Winston Washington, DC, 1977.
- [131] L. Wasserman. Asymptotic properties of nonparametric Bayesian procedures. *Practical nonparametric and semiparametric Bayesian statistics*, pages 293–304, 1998.
- [132] L. Wasserman. Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, 44(1):92–107, 2000.
- [133] J. Weese. Regularization method for nonlinear ill-posed problems. *Computer Physics Communications*, 77(3):429–440, 1993.
- [134] R.E. Whaley. Derivatives on Market Volatility. *The Journal of Derivatives*, 1(1):71–84, 1993.
- [135] C. Zhou. A Jump-Diffusion Approach to Modeling Credit Risk and Valuing Defaultable Securities. *Finance and Economic Discussion Series, The Federal Reserve Board*, 1997.